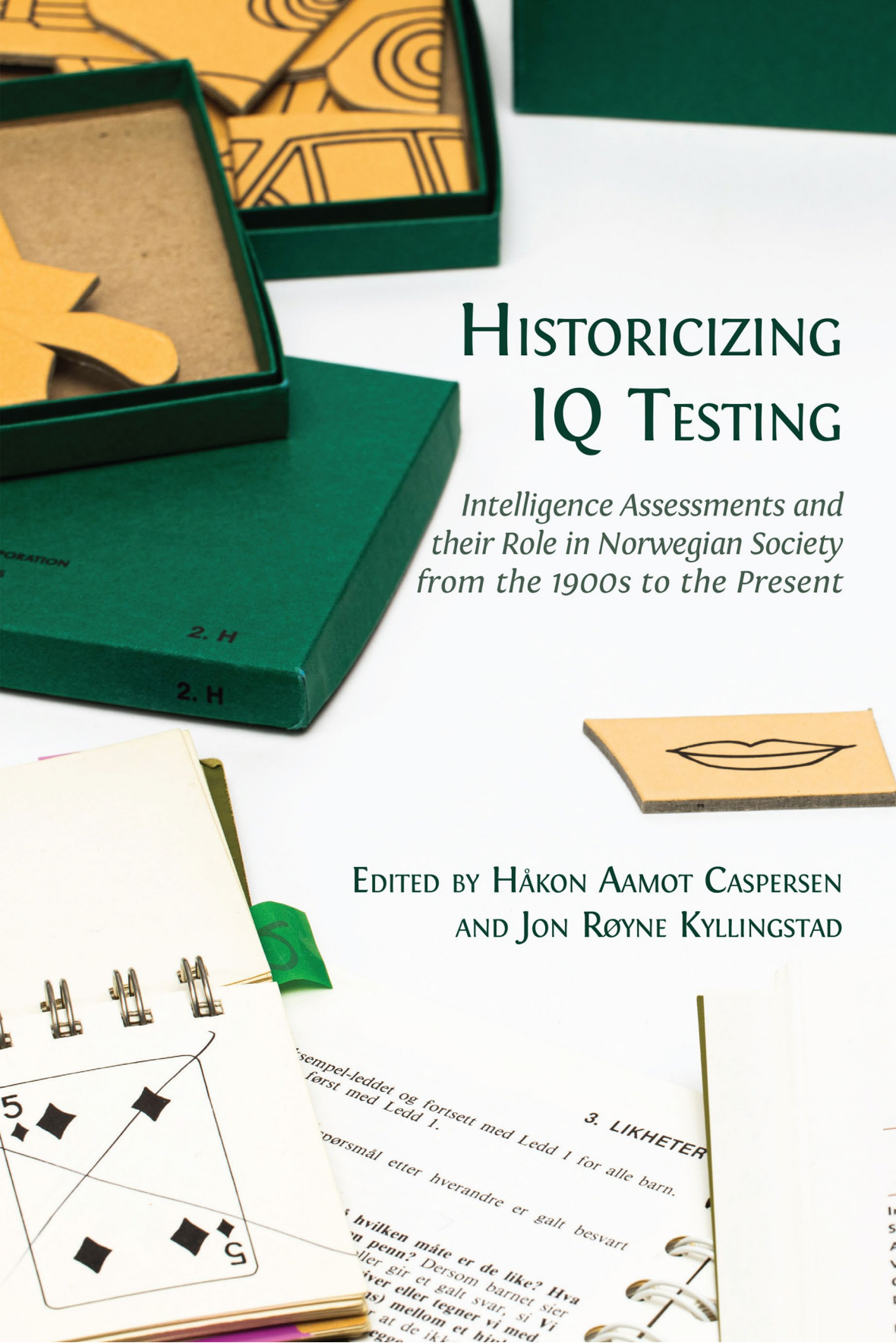




# HISTORICIZING IQ TESTING

*Intelligence Assessments and  
their Role in Norwegian Society  
from the 1900s to the Present*

EDITED BY HÅKON AAMOT CASPERSEN  
AND JON RØYNE KYLLINGSTAD

# HISTORICIZING IQ TESTING



# Historicizing IQ Testing

Intelligence Assessments and their Role in  
Norwegian Society from the 1900s to the Present

*Edited by*

*Håkon Aamot Caspersen and Jon Røyne Kyllingstad*





<https://www.openbookpublishers.com>

©2026 Håkon Aamot Caspersen, Jon Røyne Kyllingstad  
Copyright of individual chapters is maintained by the chapters' authors



This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0). This license allows you to share, copy, distribute and transmit the text; to adapt the text for non-commercial purposes of the text providing attribution is made to the authors (but not in any way that suggests that they endorse you or your use of the work). Attribution should include the following information:

Håkon Aamot Caspersen and Jon Røyne Kyllingstad (eds), *Historicizing IQ Testing: Intelligence Assessments and their Role in Norwegian Society from the 1900s to the Present*. Cambridge: Open Book Publishers, 2026, <https://doi.org/10.11647/OBP.0471>

In order to access detailed and updated information on the license, please visit, <https://www.openbookpublishers.com/product/0471#copyright>

Further details about CC BY-NC licenses are available at <https://creativecommons.org/licenses/by-nc/4.0/>

All external links were active at the time of publication unless otherwise stated and have been archived via the Internet Archive Wayback Machine at <https://archive.org/web> Updated digital material and resources associated with this volume are available at <https://www.openbookpublishers.com/product/0471#resources>

Every effort has been made to identify and contact copyright holders and any omission or error will be corrected if notification is made to the publisher.

Paperback: 978-1-80511-617-2

Hardback: 978-1-80511-618-9

PDF: 978-1-80511-619-6

HTML: 978-1-80511-621-9

EPUB: 978-1-80511-620-2

DOI: 10.11647/OBP.0471

Cover image: © Helge Brekke/MUV, UiO, CC BY-SA 4.0

Cover design: Jeevanjot Kaur Nagpal

The Cultural History Museum, University of Oslo, supported this Open Access publication.

# Table of Contents

---

|                                                                                                                              |      |
|------------------------------------------------------------------------------------------------------------------------------|------|
| Notes on Contributors                                                                                                        | vii  |
| List of Illustrations                                                                                                        | xi   |
| Preface and Acknowledgments                                                                                                  | xvii |
| Historicizing IQ Testing in Norway: Introduction                                                                             | 1    |
| 1. Between Change and Stability:<br>The Thorny Adaptation Process of the Stanford–Binet<br>Intelligence Scale in Norway      | 29   |
| 2. The Scandinavian Space of IQ Testing: Between<br>Normal and Special Education, 1918–1940                                  | 73   |
| 3. A Legal History of Intelligence, Testing, and Criminal<br>Unaccountability                                                | 101  |
| 4. IQ Testing and Sterilization in Norway, 1930 to 1960                                                                      | 131  |
| 5. Tools of the Trade:<br>Psychotechnics, IQ Testing and the Making of the<br>Psychological Profession in Norway, 1925–1947  | 159  |
| 6. From Segregation to Integration<br>and the Role of Testing in a Norwegian Educational<br>Psychology Office, 1953–1980     | 189  |
| 7. IQ Testing Today in Norway’s<br>Educational Psychological Services                                                        | 215  |
| 8. ‘Children Got Slightly Smarter with Fish for Lunch’:<br>The WPPSI Test, Randomized Trials,<br>and Optimized Kindergartens | 243  |
| 9. A Norwegian Terman–Merrill<br>and the Shelf Life of Intelligence Tests                                                    | 269  |
| 10. “Culture” and Representability in the Norwegian<br>Standardization of WISC-R                                             | 303  |

|                                                                                                                                   |     |
|-----------------------------------------------------------------------------------------------------------------------------------|-----|
| 11. 'A Violation of the Child's Integrity and of Parental Rights': The 1959 Controversy on IQ Testing of Norwegian Schoolchildren | 337 |
| 12. Tests, Metrics, and the Making of the Norwegian School, 1950–2025                                                             | 363 |
| 13. Drawing Boundaries, Building Barriers: Twentieth-Century US and Norwegian Intelligence Testing from a Comparative Perspective | 401 |
| 14. IQ Testing, Education, Reification, and "Race" in Norway's Social Democratic Welfare State                                    | 443 |
| Afterword                                                                                                                         | 459 |
| Index                                                                                                                             | 467 |

## Notes on Contributors

---

**Susanne Bauer** is professor in Science and Technology Studies (STS) at the TIK Centre for Technology, Innovation and Culture, Faculty of Social Sciences, University of Oslo. Their research into biomedical data labor, knowledge infrastructures, the work of epidemiological numbers, hybrid ecologies, and more-than-human biopolitics has been published in journals such as *Social Studies of Science*; *Science, Technology and Human Values*; *History of the Human Sciences*; *Osiris*; *Science in Context*; *Environmental Humanities*; *Somatechnics*; *Medical Anthropology Quarterly*; *Big Data and Society*.

**Håkon Caspersen** is a social anthropologist with a PhD from the University of St Andrews and currently a postdoctoral research fellow with the project *Historicizing Intelligence* at the Museum of University History/Museum of Cultural History, University of Oslo.

**Isak Lønne Emberland** is a historian with a PhD in the study of professions from Oslo Metropolitan University (OsloMet). His research focuses on the history of the psychological profession, psychological testing in the labor market, psychological sleep research and the legitimization of human sciences.

**Linda Gröning** is a professor in law, University of Bergen, and senior researcher at the Centre for Research and Education in Forensic Psychiatry and Psychology, Haukeland University Hospital, Bergen. Gröning has a research focus on criminal responsibility and the legal relevance of mental disorders and disabilities, with an emphasis on comparative and interdisciplinary perspectives. She has published extensively on this topic.

**Per Haave** (1958–2024) was a recognised historian of medicine and psychiatry in Norway. Haave was a Norwegian Government Grant

holder and researcher at the Institute for Health and Society, University of Oslo. He published extensively on topics such as the history of psychiatry, the medical profession, the welfare state, and eugenics and involuntary sterilization.

**Kim Helsvig**, historian and professor at Oslo Metropolitan University, has primarily worked and published on modern intellectual and educational history, including monographs about Norwegian educational science and policy, the University of Oslo, The Norwegian Academy of Science and Letters, and the Norwegian Ministry of Education and Research.

**Brit Marie Hovland** is associate professor, VID Specialized University, Oslo. Hovland's expertise lies within the history of education, historiography, curriculum, didactics and pedagogy. Her main research topics has been national identities and nation-building, historiography, curriculum, the contextual changing epistemology and narratology of school subjects, child-centered education, and more. Currently she focuses on interwar progressive education and curriculum reforms, including Sagene exploratory school and experimental psychology, IQ and tests. She is the Norwegian leader of the international NFR research project *Transloyalties in History and Citizenship Education*. Hovland is scientific expert on the Swedish Research Council.

**Jon Røyne Kyllingstad** is a historian and associate professor at the University of Oslo, Museum of University History/Museum of Cultural History, where he is the leader of the research project *Historicizing Intelligence*, which this book is based upon. He is a specialist in the history of science and the history of academic institutions with a focus on Norway. He was previously head conservator at the Norwegian Museum of Technology. His last book *Rase: en vitenskapshistorie* [Race: a history of a science] sums up two decades of work on changing ideas about race, ethnicity and the nation, within physical anthropology, genetics, and humanities disciplines such as archaeology and history in Norway. Similar topics were also addressed in *Measuring the Master Race*, published by Open Book Publishers in 2014.

**Ageliki Lefkaditou** is a historian and philosopher of science with PhDs from the University of Leeds (UK) and the Aristotle University of Thessaloniki (Greece). Her research focuses on the history of the human

and biological sciences from the mid-nineteenth to the mid-twentieth century. She is currently investigating the history of intelligence testing in Norway and has previously published on racial science and its connections to nationalism. Ageliki is particularly interested in the intersections of science, art, and politics, and issues related to nature/culture. She has curated several science museum exhibitions and is an award-winning curator and documentary film producer.

**Kari Ludvigsen**, is a professor at the Western Norway University of Applied Sciences, Faculty of Teacher Education, Culture and Sports, Bergen, Norway. Ludvigsen is a political scientist teaching and researching topics regarding education and welfare state expertise, policies and services. She has published several works on the role of psychology, psychiatry and pedagogy in categorizing children on the borders between mental health services, social care and education, with an emphasis on the period from 1920–1960.

**Annette Mülberger** is chair at the *Theory & History of Psychology* unit at the University of Groningen. Her research deals with the history of psychology, criminology, and education, and, specifically, with IQ testing. In 2014 she coordinated a monographic issue (*History of Psychology*, vol. 17.3) on mental testing in different local contexts, and in 2020 her entry on the intelligence test in the *Oxford Research Encyclopedias for History of Psychology* appeared.

**Gard Paulsen** is a researcher at the Department of Philosophy, Classics, Art History and the History of Ideas, University of Oslo. He has published on the history of cybernetics, telecommunications, maritime concerns, risk and regulation.

**Jim W. Porter** earned his PhD in history from Michigan State University. He then served as a postdoctoral fellow and researcher at the Hugo Valentin Centre, Uppsala University, and subsequently as a research fellow at the Science History Institute and the American Philosophical Society in Philadelphia. His first book, *Safeguarding Whiteness*, is in revision with Johns Hopkins University Press. His research explores the rapid implementation of ‘intelligence’ policy (and related standardized testing practices) in the years immediately following the 1954 Brown v. Board mandate to desegregate public schools.

**Svein Atle Skålevåg**, professor in the department of archaeology, history, cultural studies and religion, University of Bergen. Skålevåg teaches history of science at the University of Bergen and has for many years been researching the history of psychiatry and in particular of forensic science. He has written *Utilregnelighet. En historie om rett og medisin*, a book length study of the history of criminal insanity in forensic psychiatry and law in Norway from the eighteenth to the twenty-first centuries.

**Fredrik W. Thue** is a professor in the history and theory of professions at Oslo Metropolitan University. He has published on university history, the history of the social sciences, educational history, and the history of historiography, and is currently writing a monograph on the Lutheran pastor as a historical predecessor and model of such modern “caring professions” as teachers, nurses, and social workers.

**Emma Vikström** is a postdoctoral fellow in history at Umeå University, Sweden. In her dissertation, *The Creation of the New Human*, she analyzed the relationship between education and eugenics in the work of Swedish progressive educator Ellen Key. Vikström’s overall research interests lie in the field of inclusion and exclusion in historical educational contexts, and her current postdoctoral project focuses on the relationship between race biology and gender in Swedish educational debates during the first half of the twentieth century.

**Christian Ydesen** is Chair of History of Education and Education Policy Analysis at the Department of Education, University of Zürich, Switzerland, and an Honorary Research Fellow at the Department of Education, University of Oxford, United Kingdom. He has held visiting researcher positions at the University of Edinburgh (2008–2009, 2016), the University of Birmingham (2013), the University of Oxford (2019), and the University of Milan (2021, 2024). His research explores international organizations, the global governance of education, educational testing and assessment, as well as accountability and diversity in education from historical, comparative, and international perspectives. He has published extensively in these areas, contributing numerous book chapters and journal articles. Since 2025, he has served as Editor-in-Chief of the *European Educational Research Journal*.

# List of Illustrations

---

- Fig. 1.1      Portrait of psychiatrist Johan Lofthus (c.1935). He worked with the adaptation of the Stanford–Binet scale in Norway after he obtained the position of school psychiatrist in Oslo in 1923. He was a staunch ambassador for the use of intelligence tests in schools, prisons and mental institutions and collaborated closely with school staff, doctors and psychologists to advance this cause. After the publication of his standard, known as *Osloprøvene* [Oslo Tests], he left his position. Photographer unknown/Oslo Museum, CC0 1.0. p. 50
- Fig. 2.1      Helga Eng (1876–1966) was the first person to earn a doctorate in psychology in Norway (1913). In 1938, she became the inaugural professor of pedagogy at The Pedagogical Research Institute, Det Kongelige Frederiks Universitet in Oslo. She was a pioneering educationalist, pedagogue, and educational psychologist. Her many books range from test psychology to art pedagogy and they were translated into several languages. Her wide-ranging influence and academic position situate her as an international pioneer within—and between—pedagogy and test psychology in the period the two disciplines diverged. Photographer unknown/Oslo Museum, CC0 1.0. p. 86
- Fig. 3.1      Caricature of Professor Ragnar Vogt (1870–1943). oral degree in 1901 on research in experimental psychology, and in 1915 Vogt became the first professor of psychiatry in Norway. He sat for many years as a member of the Board of forensic psychiatry. Photo: Norsk Teknisk Museum, CC BY-NC-SA. p. 111



- Fig. 3.2 Paul Winge (1857–1920) studied medicine in Oslo, and earned his doctoral degree in 1896 on a dissertation titled ‘The forensic medical examination of the insane criminal’. He was employed as a police and prison physician from 1893 to his death in 1920. In this role he encountered and examined a large number of criminals, probably more than any other Norwegian physician of his time. He was instrumental in establishing a Norwegian Board of Forensic Medicine, and sat as one of its psychiatric members for twenty years. Photographer L. Forbech 1901/ National Library of Norway, PDM 1.0. p. 111
- Fig. 4.1 Table showing an overview of the number of men and women diagnosed as “mental defectives”, 1945–1954, compiled by the authors. For source information see Karl Evang, ‘Sterilisering etter lov av 1. juni 1934 om adgang til sterilisering m.v.’ Bilag til Tidsskrift for Den norske lægeforening, nr. 2, 1956: 16 (Table 2), 31 (Table 10), 39 (Table 12). p. 138
- Fig. 4.2 Karl Evang (1902–1981), Norwegian General Director of Health 1938–1972, he was involved in the establishment and leadership of the World Health Organization. During the interwar years, he became known to the Norwegian public as a radical socialist, an eager advocate for sexual education, family planning and the legalization of abortion. In the pamphlet *Rasepolitikk og reaksjon* [Racial Politics and Reaction] (1934), he strongly criticized racial thinking and a racist and classist style of eugenics, while advocating a non-racial and class-neutral style of eugenics, as part of a scientifically organized socialist society. Photographer Rigmor Dahl Delphin/Oslo Museum, CC0 1.0. p. 140
- Fig. 4.3 Augusta Rasmussen (1895–1979) was, in 1933, the second woman in Norway to become a specialist in psychiatry. The same year she succeeded Johan Lofthus in the position of school psychiatrist in Oslo (1933–1965). In 1932 she published the award-winning study ‘The impact of sexual assaults on children under 14 years of age for the development of mental disorders and character anomalies’. p. 144

Here, she claimed that most of the children she had examined had suffered no bodily or mental injury from the sexual assault. In 1947, she published a study based on the IQ testing of 310 Norwegian women who had had sexual relationships with German occupation soldiers, and claimed to have found that nearly all of them were more or less intellectually disabled. In retrospect, her research has faced strong criticism, but in her time, she was a recognized professional and was often used as an expert witness in the courtroom. Photographer unknown/Arbeiderbladet/Arbeiderbevegelsens archive and library. All rights reserved.

- |          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |        |
|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| Fig. 4.4 | Nic Waal (1905–1960) was a psychiatrist, one of Norway’s first psychoanalysts, a pioneer in child psychiatry, and, from 1951, founder and director of the Nic Waals Institutt for barne-og ungdomspsykiatri [The Nic Waal Institute of Child and Adolescent Psychiatry], which is still a leading institution in its field in Norway. Like Karl Evang, she was a member of the left-wing radical student group Mot Dag in her youth. In 1943, she was imprisoned by the German occupying authorities for her participation in the resistance movement. Photo: Arbeiderbladet 1958/Arbeiderbevegelsens arkiv og bibliotek, CC BY NC ND 4.0. | p. 151 |
| Fig. 5.1 | Harald Krabbe Schjelderup (1895–1974), a Norwegian psychologist and one of the authors of the white paper establishing psychology as a stream of professional education at the University of Oslo in 1948. Photographer unknown/Museum of University History, KHM, University of Oslo. All rights reserved.                                                                                                                                                                                                                                                                                                                                | p. 163 |
| Fig. 5.2 | Anathon Aall (1867–1943), Norwegian philosopher and psychologist. Established experimental psychology at the University of Oslo. Photographer Fritz Möller, 1904/National Library of Norway, PDM 1.0 (Public domain).                                                                                                                                                                                                                                                                                                                                                                                                                      | p. 168 |
| Fig. 5.3 | Helga Eng working on psychotechnical tests, photo taken some time between 1925–1935. Photographer unknown/Oslo Museum, CC0 1.0.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | p. 172 |

- Fig. 6.1 Åse Gruda Skard (1905–85) was a Norwegian psychologist who pioneered mental hygiene for schoolchildren, teachers, and parents from the 1930s. From her position in psychology at the University of Oslo, she was active in research and in international associations as well as a popular educator in childhood and upbringing. Photographer Bachrach 1941/National Library of Norway, PDM 1.0 (Public domain). p. 194
- Fig. 7.1 David Wechsler (1896–1981) not only created highly influential IQ tests, he also experimented with other physiological and psychological measurements. The identity of the exact test being administered in this photograph is unknown. Photographer unknown, 1955/New York University, School of Medicine, CC BY 4.0. p. 236
- Fig. 8.1 WPPSI-III Test box; WPPSI-III Manual; example from ‘Concept Grouping’ part of the performance IQ test (picture concepts, supplemental sections). Photographs by the author. CC BY-NC. p. 246
- Fig. 8.2 Test examples: Fish puzzles in WPPSI-III test (materials from the Swedish and Norwegian version). Photographs by the author. CC BY-NC. p. 247
- Fig. 8.3 Design box. Organizing the intervention and the study population (plugging Wechsler tests into a nutrition trial). Published in Jannike Øyen, Ingerid Kvestad, Lisa Kolden Midtbø, et al. ‘Fatty Fish Intake and Cognitive Function: FINS-KIDS, a Randomized Controlled Trial in Preschool Children’, *BMC Medicine*, 16.1 (2018). CC BY 4.0. p. 251
- Figs. 8.4a and 8.4b Tables including Wechsler raw scores; results from WPPSI tests, first published in Øyen et al, 2018; Kvestad et al, 2018. CC BY 4.0. p. 256
- Fig. 9.1 Helga Eng’s own Terman Merrill 1937 edition. Photographer Helge Brekke/Museum of University History (MUV), KHM, University of Oslo. All rights reserved. p. 277
- Fig. 9.2 Material from the standardisation of the Norwegian adaption of the 1937 Terman-Merrill test. Photographer Helge Brekke/Museum of University History (MUV), KHM, University of Oslo. All rights reserved. p. 285

|           |                                                                                                                                                                                                                                                                                                                                                       |        |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| Fig. 9.3  | The Norwegian assembly of a Terman-Merrill test. Photographer Helge Brekke/Museum of University History (MUV), KHM University of Oslo. All rights reserved.                                                                                                                                                                                           | p. 288 |
| Fig. 10.1 | WISC-R test kit. Photo: Helge Brekke/Museum of University History, KHM. University of Oslo. All rights reserved.                                                                                                                                                                                                                                      | p. 313 |
| Fig. 10.2 | WISC-R block design. Photo by Jon R. Kyllingstad. CC BY-NC.                                                                                                                                                                                                                                                                                           | p. 314 |
| Fig. 10.3 | WISC-R picture completion. Photo by Jon R. Kyllingstad. CC BY-NC.                                                                                                                                                                                                                                                                                     | p. 314 |
| Fig. 10.4 | WISC (1949) Picture arrangement. Photo: Helge Brekke/Museum of University History, KHM, University of Oslo. All rights reserved.                                                                                                                                                                                                                      | p. 315 |
| Fig. 10.5 | Even if this playing card was not included in the Norwegian WISC-R version, the test kit was imported unmodified. Photo: Jon R. Kyllingstad. CC BY-NC.                                                                                                                                                                                                | p. 322 |
| Fig. 12.1 | The psychologist and pedagogue Helga Eng (1875–1966) was a pioneer in developing psychometrics in Norway, first as leader of the Psychotechnical Institute from 1925, and, from 1938, as the first professor in pedagogy at the University of Oslo. Photographer unknown, Museum of University History, KHM, University of Oslo. All rights reserved. | p. 365 |



## Preface and Acknowledgments

---

This book has come to life as part of the research project *Historicizing Intelligence: Tests, Metrics, and the Shaping of Contemporary Society*, funded by the Norwegian Research Council and based at the Museum of University and Science History (MUV), which is part of the Cultural History Museum, University of Oslo, Norway. The project emanates from MUV's mandate to manage, disseminate, and research the historical legacy of the university. One of the sources of inspiration for the project has been an abandoned assemblage of old tests and archive materials that belonged to the Pedagogical Research Institute. The material was found in the institute's present incarnation—the Institute of Pedagogy, part of the Faculty of Education—in the cellars of a building named after Norway's first professor of education, Helga Eng, whose name echoes in many of the following chapters. With the aid of work by Solveig Siem, this array of objects was categorized, catalogued, and transformed into a museum collection that can now be accessed through the University Library of Oslo's special collection portal.<sup>1</sup> Literally opening boxes has, in fact, been key for many of the contributors in making sense of and getting to know these texts, with their histories of kinship, journeys, development, and material make-up.

While rooted in the history of science, the contributors include historians of education, psychology, psychiatry, eugenics, and criminal law, as well as political science, social anthropology, STS, and law. By bringing together different historical traditions, and disciplinary perspectives and angles on IQ testing—and by putting concrete, in-depth historical and contemporary empirical case studies from Norway side by side—we hope to lay the ground for future integrative comparison, to help clarify differences and continuities across contexts, places, and

---

1 The special collection portal can be found here: <https://www.ub.uio.no/om/prosjekter/spesialsamlinger-og-databaser/samlinger/IPED-i-alvin.html>

times. We see this as an important step, both in creating new knowledge around how these tests have historically functioned, and in recalibrating our understanding of IQ testing and the contemporary testing practices that are present in Norway and the world today.

The chapters emerge from discussions and presentations during three two-day workshops in Oslo over a period of two years. In addition to this volume's contributors, the workshops were attended by members of the project's Advisory Board. In particular, we thank Annette Mülberger, Thom Axelsson, and Christian Lundahl for their always insightful but also critical comments during our discussions, which have inspired and influenced the development of all the chapters for the better. We also greatly appreciated the constructive comments by Sasha Bergström-Katz at the final workshop. We are indebted to our former project coordinator, Anne-Marit Hessevik, for organizing the events—and all of us—and for her contributions to the discussion and planning of the volume.

We are also grateful for the useful and critical commentary of Eric Turkheimer, as well as one other anonymous peer reviewer; and to Alessandra Tosi, Lucy Barnes, and the rest of the team at Open Book Publishers for their support in making this book a reality. The Cultural History Museum, University of Oslo, generously contributed towards the production costs. Kriti Bajaj provided invaluable help with the final copy edit of the manuscript.

Lastly, our thanks go to all of our contributors for your patience and dedication throughout the book's conception. It has been a pleasure working with you, and we have learned a lot.

On 18 January 2024, Per Haave, our beloved project member, colleague, and friend, abruptly passed away only two weeks before the final workshop, and the day after completing his chapter draft for this volume. This book is dedicated to his memory.

Jon Røyne Kyllingstad, Håkon Aamot Caspersen

# Historicizing IQ Testing in Norway: Introduction

*Håkon Aamot Caspersen and Jon Røyne  
Kyllingstad*

---

The power of Intelligence Quotient (IQ) tests rests on their legitimacy as scientific instruments, and the fact that they are developed, controlled, and used by experts with legal, public, and academic authority. Intelligence testing, as it developed historically, has offered specific ways of visualizing and codifying individual performance into ‘writing and numbers, quotients, scores, profiles’, thus making ‘the individual knowable, calculable and administrable’ in new ways.<sup>1</sup> But there has rarely been a consensus about exactly what IQ tests measure, or clear agreements about their fairness (or lack thereof) towards different ethnically, socially, or racially defined groups.

IQ tests are deeply institutionalized in Norwegian society today, forming part of many official assessment, diagnostic, and selection procedures, often in cases that involve the allocation of specific rights and resources. They contribute to practices of constructing and categorizing the normal and abnormal, and have existential, pedagogical, social, legal, and economic consequences, both individually and societally.

Given their influence, it is crucial to develop a critical understanding of intelligence tests and the practices of testing. This book contributes towards realizing such an objective by historicizing IQ testing, asking questions about how these practices came into being, and how they achieved the functions they currently have in our society. Our aim is

---

1 Nikolas Rose, *Governing the Soul: The Shaping of the Private Self* (London: Free Association Books, 2<sup>nd</sup> ed. 1999), p. 143.



to open up historical and contemporary understandings of IQ testing, inviting a closer look into this particular socio-technological practice and its associated disciplinary boxing practices (to borrow phrasing from Susanne Bauer in chapter 8).

Throughout the chapters, themes that recur include the importance of distinguishing between how IQ testing and intelligence have been explained, and how IQ tests have been used—that is, the practical operationalization of IQ and intelligence. An IQ test, in this sense, is multi-faceted—a practical clinical and bureaucratic tool, a scientific method, as well as a communicative device that spans national borders, societal sectors, and disciplines. It is often in its practical applications, and in conversations across disciplinary fields, that one sees the paradoxical reification of IQ as an objective and frequently black-boxed technological measure in society, despite more nuanced understanding among the holders of test expertise. It is such themes, relationships, apparent paradoxes, and more, that we examine in this volume.

## The Significance of a Test Score: An Example from Norway

In 2017, journalistic investigations brought media attention to multiple cases of misdiagnosis of intellectual disability, following the use of one specific IQ test—the third edition of the Wechsler Intelligence Scale for Children (WISC-III).<sup>2</sup> A low score on an IQ test is a key criterion for diagnosing intellectual disability, and the WISC tests are the most commonly used IQ tests for children in Norway and internationally. The investigation showed that recommendations of sterilization and cases of forced adoption grounded in mental disability misdiagnoses partly followed from a low IQ score caused by a flaw in the Norwegian version of the test. This testifies to the far-ranging consequences an IQ test may have. A disability diagnosis will inform and significantly shape lives, and people's legal rights.

---

2 See Asbjørn Øyhovden, Connie Bentzrud and Olav T. Hustad Wold, 'Metoderapport SKUP 2017: IQ Testing og Diagnosehelvete', in *SKUP: Stiftelsen for en kritisk undersøkende presse* (2017). Examples from the news coverage: Asbjørn Øyhovden and Connie Bentzrud, 'Slo alarm om IQ: Barn og unge kan ha blitt stempet som utviklingshemmede etter feil' (TV2.no, 26.02.2017), Asbjørn Øyhovden, 'Strid om ansvaret for at barn og unge kan ha fått feil IQ-diagnose' (TV2.no, 01.03.2017), Asbjørn Øyhovden and Connie Bentzrud, 'Helseministeren vil rydde opp i IQ-testing' (TV2.no, 28.02.2017).

As early as 2005, two years after the introduction of the Norwegian version, experienced psychologists had warned of a mismatch between their own clinical judgment and scores produced by this test.<sup>3</sup> Despite warnings and advice of caution in interpreting the results, the test was widely used. The mismatch, later confirmed by frequent cases of misdiagnosis, indicates too much confidence in test results, and their institutionalized position as indispensable tools in diagnostic work. It was later documented that the Norwegian version of the test consistently indicated lower scores than its predecessor and comparable test batteries. An analogy employed to characterize the problem was that the scale showed the wrong ‘weight’.<sup>4</sup> The critique did not question IQ as a valid measurement of ability, but located the problem in the technical process of adapting this specific test for use in Norway. That, in turn, produced an “error” in the IQ scale. However, these technical issues were related to the economic, legal, institutional, and societal framework within which this work was undertaken. The development of the Norwegian version was not influenced only by scientific considerations. The state rejected a request from the Norwegian Psychological Association for the funding of a Norwegian adaptation. Instead, the work was carried out by a newly established private company that aimed to finance the project through test sales. For this endeavor to be economically viable, taking the smaller scale of the Norwegian test market into consideration, they had to keep costs down.<sup>5</sup>

As several of the following chapters show, issues of adaptation, translation, and statistical standardization of a test—including the social work necessary to legitimate them as scientific tools of measurement—are intimately tied in with different historical trajectories and social and economic concerns that have influenced the development of IQ tests. Far from infallible instruments, IQ tests and corresponding ideas about intelligence, like other technologies of measurement, are constructed, and need to be understood as historically contingent ‘scientific objects’.<sup>6</sup>

---

3 As reported by, for instance, Nina Strand, ‘WISC-III skaper uro’ (Psykologitidskriftet.no, 01.06.2025).

4 Thor Herman Andreassen and Øyvind Fallmyr, ‘Kan vi leve med manglende norsk normering på WISC, WAIS og WPPSI?’ (Psykologitidskriftet.no, 01.05.2006).

5 *Tidsskrift for Norsk psykologforening*, 40.1 (2003), 52–53; 39.2 (2002), 130–131; and 40 Nr. 8 (2003), 704–705. *WISC-III Norsk versjon Administrasjonsveiledning* (Oslo: Assessio Norge 2003), pp. 5–7.

6 Lorraine Daston, ‘Introduction: The Coming into Being of Scientific Objects’, in

This entails discussions that go beyond internal epistemic critique or explorations of the validity of a test in relation to certain scientific criteria. The starting point of this book is that a test is a cultural object that participates in producing the very thing it purports to measure.

## The Binet–Simon and the Wechsler Test Families: A Brief Introductory Sketch

The WISC-III belonged to one of two “families” of intelligence tests—the Binet–Simon test and its successors, and the Wechsler tests—that have been central to the international history of IQ, and which play key roles in this book. The history of psychological testing began in the nineteenth century, with the English polymath Francis Galton as a key contributor. Galton coined the term ‘eugenics’, and advocated the improvement of humankind through interventions in the biological reproduction of humans to ensure that the talented had more children than the less talented. Galton saw intelligence as closely linked to sensory capacities and speed of reaction, and developed a method for measuring intelligence based on non-verbal, sensor-motoric tests. It is, however, common and reasonable to localize the birth of IQ testing proper to Paris at the beginning of the twentieth century, arising from the work of the psychologist Alfred Binet and physician Théodore Simon. This was born partly in response to the need for a tool that the French school system could use to identify children who required special education measures. Binet and Simon developed an intelligence scale with standardized scoring procedures, published in 1908, which involved the individual child responding to a set of tasks, and test results were measured against the average test result among Parisian children of different age groups. By comparing these scores, the child’s “mental age” was calculated. This opened the way for the understanding of intelligence as a uniform entity unevenly distributed between individuals. The first instances of intelligence testing in Norway were done using more or less direct translations of the French Binet–Simon test.

While gaining a limited foothold in France, the Binet–Simon test had

---

*Biographies of Scientific Objects*, ed. by Lorraine Daston (Chicago: University of Chicago Press, 2000). See also F. A. Hanson, *Testing, Testing: Social Consequences of an Examined Life* (Berkeley: University of California Press, 1993).

its greatest success in the USA.<sup>7</sup> It was first translated and propagated as an instrument in the scientific study of the “feeble-minded” by psychologist Henry H. Goddard, the director of the Vineland Training School for Feeble-minded Boys and Girls. In 1916, his colleague Lewis Terman transformed, adapted, and standardized the test with scoring norms adjusted to an American population. It was in this new US version—the so-called Stanford–Binet—that the Intelligence Quotient was introduced. The IQ, originally developed by German psychologist William Stern, was calculated by dividing mental age by biological age and multiplying by one hundred. In contrast to Binet, who had been agnostic about the effect of nature and nurture on measured intelligence age, Terman largely portrayed IQ as a direct measure of inherited ‘general intelligence’.<sup>8</sup>

In Britain, Galton’s early influence was carried forth by figures like Charles Spearman and Cyril Burt. Spearman observed that children’s performance across seemingly unrelated school subjects was statistically correlated, and in 1904, he put forward the theory that these correlations were produced by a single, underlying, general intelligence. Cyril Burt—who eventually succeeded Spearman as professor of psychology at University College London in 1932—was appointed as psychologist to the London County Council in 1913. He adapted the Binet–Simon and created his own tests while promoting Galton’s hereditarian stance on intelligence. From his institutional position and his strong interest in the psychology of individual differences, with support from Spearman, he played a significant role in institutionalizing both intelligence testing and educational psychology as scientific research programmes. Alongside other psychologists and educational thinkers of the time, including in Norway, Burt saw IQ testing as a tool that offered the combination of meritocratic ideas of educational selection, a scientific passion for measurement, and a commitment to child-centred education.<sup>9</sup>

---

7 John Carson provides an excellent in-depth account of these comparative developments in *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics 1750–1940* (Princeton: Princeton University Press, 2007).

8 See Jim Porter, chapter 13, for a further expansion on these developments, building on previous historical accounts.

9 See Adrian Wooldridge, *Measuring the Mind: Education and Psychology in England, c.1860–1990* (Cambridge: Cambridge University Press, 1994) and Gillian Sutherland, *Ability, Merit and Measurement: Mental Testing and English Education 1880–1940* (Oxford: Clarendon, 1983) for good overviews of ability measurement and educational

Despite multiple international adaptations of its French original, it is the US-based Stanford–Binet that perhaps can be counted as the most influential, and it became the model for many other tests widespread in the education system, healthcare, and armed forces in the US and beyond. It was also the first intelligence test that was successfully adapted and normed for and made mainstream in Norway in 1931. By ‘normed’, we here refer to how the calculation procedures for converting raw test scores into an IQ score were adjusted, based on the average distribution of test scores among Norwegian school children. It was only after World War Two that the Binet tests were challenged as the most influential way to measure intelligence, through the introduction of the Wechsler tests.

The first Wechsler test was launched by psychologist David Wechsler in 1939.<sup>10</sup> He did away with the practice of calculating the IQ score by dividing mental age by biological age, and introduced the method that dominates today. According to this method, an IQ score is calculated by statistically comparing the test scores of an individual with the average score of a “representative” group that stands for the population and age group that the individual belongs to. The scale is constructed on the principle that general intelligence, as measured by IQ, is normally distributed within the total population, forming a bell curve. An IQ score of 100 is set as the median, and 68 percent of the population scores between 85 and 115, defined as the normal range.

The Wechsler tests come in numerous versions for pre-school children, school-age children and adolescents, and adults, and the latest versions have been translated and adapted for use in nations around the world. Since at least the 1970s, they have been the most used intelligence tests in Norway. They are standard assessment tools used by the Educational Psychological Service (EPS) and in psychiatric healthcare. It is through being certified to use this test that students in psychology, educational psychology, and special needs education learn to use intelligence tests and psychometric tests more generally.

The Binet–Simon and Wechsler test families are only a few among a

---

psychology as it developed in the UK. While interesting parallels and differences certainly exist between the UK and Norwegian cases, it is the standard history from the US that takes centre stage in the following pages, as it was US versions of tests that dominated in the work of Norwegian historical translations.

10 See David Wechsler, *The Measurement of Adult Intelligence* (Baltimore: Williams & Wilkins, 1939).

huge number of differently composed IQ tests. These varied tests and test batteries are like complex, unsystematic kinship systems. Not only do different versions develop and build upon each other, but as they develop, they borrow and combine new elements and subtests from a range of other tests, whether IQ tests or otherwise.<sup>11</sup> The basic structure of a Wechsler test, however, remains the same as its predecessors. They are scored according to the number of right or wrong answers to questions of increasing difficulty on different subtests, understood today to relate to a set of cognitive factors associated with intelligence.

Intelligence tests are tests by proxy. It is not “intelligence” in and of itself that is measured, but purported—and historically changing—representations of it, through subtests that stand for, or index, underlying cognitive factors. Many IQ tests today will provide not only a full-scale IQ score, but also a cognitive profile meant to indicate specific strengths and difficulties.

## The Historiography of IQ Testing as Viewed from the Periphery

There is extensive literature on IQ tests, their historical development, and their role in the discipline of psychology and broader society. While a comprehensive review of the historiography of IQ testing is beyond our scope here, excellent overviews can be found elsewhere, and relevant literature is also referred to in the individual contributions to this volume.<sup>12</sup> It is, however, important to note that much of the historical

---

11 As seen, for example, in Corwin Boake, ‘From the Binet–Simon to the Wechsler–Bellevue: Tracing the History of Intelligence Testing’, *Journal of Clinical and Experimental Neuropsychology*, 24.3 (2002); Jacy L. Young, ‘Test or Toy? Materiality and the Measurement of Infant Intelligence’, *History of Psychology*, 18.2 (2015).

12 See Anette Mülberger Rogele, ‘Biographies of a Scientific Subject: The Intelligence Test’, in *Oxford Research Encyclopedias: Psychology* (Oxford: Oxford University Press, 2020) for an excellent, concise and recent overview. See also Anette Mülberger, ‘The need for contextual approaches to the history of mental testing’, *History of Psychology*, 17.3 (2014), 177–186 and the special issue of the *History of Psychology* it introduces. John Carson, in turn, provides a thorough comparative analysis of intelligence testing in the US and France in the first part of the twentieth century. For good accounts of their UK expansion see also Gillian Sutherland, *Ability, Merit and Measurement: Mental Testing and English Education 1880–1940* (Oxford: Clarendon, 1983) and Wooldridge, *Measuring the Mind*. For a disciplinary “textbook account”, see, for instance, John D. Wasserman’s ‘A History of Intelligence Assessment: The Unfinished Tapestry’, in

literature is framed by critical discussions about IQ testing and its societal consequences. On the one hand, there are highly critical narratives that emphasize IQ tests as scientifically dubious and socially harmful.<sup>13</sup> This critical stance, that often roots itself in a Foucauldian critique of power, highlights the tests as tools for sorting and governing individuals and populations, and often points to the historical connections between IQ testing, scientific racism, and eugenics. On the other hand, within the field of psychometrics and in the disciplinary histories of psychology and IQ testing textbooks, the development of intelligence testing is often portrayed as a history of steady scientific progress, with “bad” testing practice largely presented as something of the past.

A uniting feature of many historical accounts, however, is that the history of IQ testing has often been narrated as a story where IQ tests grow, travel, expand, and conquer the world. As the historian of psychology Annette Mülberger has noted, this is a story with a lot of repetition and overlap, perhaps even showing ‘a dangerous tendency toward creating [and recreating] myths’.<sup>14</sup> Furthermore, there is a propensity on both sides—critics and supporters—to overemphasize and even caricature the historical reaches of intelligence testing and the intentions of its cast of actors. The strong historiographic focus on intelligence testing and its significance in the development and cooperative triage of the psychological sciences, education, and governance risks blindness to important alternative influences. It may ignore the influence of other aspects of psychology and tests that historically have competed with IQ testing, or that came into being in cases where IQ testing was found

---

*Contemporary Intellectual Assessment*, ed. by Dawn P. Flanagan and Erin M. McDonough (Guilford, 2022). Sattler’s historical account stands as an interesting comparative example of this genre, in *Assessment of Children: Cognitive Foundations and Applications* (J. M. Sattler Publishing Inc. 6<sup>th</sup> ed., 2018). Stephen J. Gould’s well-known *Mismeasure of Man* (New York: W. W. Norton & Co, 1996) stands as a classic in the tradition of test critique, with *The Bell Curve: Intelligence and Class Structure in American Life* by Richard J. Herrnstein and Charles Murray (New York: Free Press, 1994) straddling the opposite side. For a recent insightful practice-based investigation of the social and material life of historical intelligence tests see Sasha Bergstrom-Katz, *On Intelligence Tests: Psychological Objects and Their Subjects*. (PhD thesis, Birkbeck, University of London, 2022). See also Kurt Danziger, *Constructing the Subject: Historical Origins of Psychological Research* (Cambridge: Cambridge University Press, 1990).

13 See, for instance, Tom Axelsson, ‘Att konstruera begåvning—Debatten om IQ (Malmö University electronic publishing, 2012) for an overview of central voices in the critical debate on IQ.

14 Mülberger, ‘The need for contextual approaches’, p. 181

lacking. This is forcefully shown by Bonnie Evans in her account of the changing form and phenomenal rise and transformation of autism diagnoses.<sup>15</sup>

Another aspect that existing historical narratives around intelligence testing and IQ measurement often have in common is a US and Anglo-Saxon focus that is additionally characterized by what John Carson describes as a ‘meta-narrative’, where standardization processes, such as the rise of the IQ scale, are framed as an ‘imperial story, in which a metropolitan institution defines [...] a standard, and then demands that provincials everywhere adopt that standard’. Carson points to an alternative approach which ‘emphasizes less the co-optation of the periphery by the center, and more the adoption, adaptation, and repurposing of artifacts and ideas as they circulate between putative metropolises and provinces.’<sup>16</sup> This is our starting point—that a proper historical understanding of these tests necessitates an in-depth empirical focus, which can capture the diversity of the different local and transnational constellations that IQ testing has been and is part of.

This book is positioned in relation to and against “imperialist” narratives of the development and use of IQ tests. It is inspired by recent efforts to de-center this story, as called for by Carson and Mülberger, and to study the multiple historical trajectories in a range of places and nations, including France, Brazil, Spain, Italy, Denmark, Sweden, and the Soviet Union.<sup>17</sup>

Although several accounts explore important transnational and/or

---

15 Bonnie Evans, *The Metamorphosis of Autism: A History of Child Development in Britain* (Manchester: Manchester University Press, 2017).

16 John Carson, ‘Mental Testing in the Early Twentieth Century: Internationalizing the Mental Testing Story’, *History of Psychology*, 17.3 (2014), p. 254. See also Mülberger, ‘The need for contextual approaches’; Mülberger Rogele, ‘Biographies’.

17 For France, see Carson, *The Measure of Merit*; see also the individual contributions to Mülberger’s edited special issue ‘Mental Testing after 1905: Uses in Different Local Contexts’, *History of Psychology*, 17.3 (2014) for examples from Spain, Italy, Brazil, and more. For Denmark, see Christian Ydesen, Bjørn Hamre, and Karen Egedal Andreassen, ‘Differentiation of Students in the Early Danish Welfare State: Professional Entanglements Between Educational Psychologists and Psychiatrists’, *Nordic Journal of Educational History*, 5.1 (2018). For Sweden, Tom Axelsson, *Rätt elev i rätt klass. Skola, begåvning och styrning 1910–1950* (Avh. Linköpings Universitet, 2007). For the Soviet Union, see Irina Leopoldoff, ‘A psychology for pedagogy: Intelligence testing in USSR in the 1920s’, *History of Psychology*, 17.3 (2014); Andy Byford, ‘The Mental Test as a Boundary Object in Early-20-Century Russian Child Science’, *History of the Human Sciences*, 27.4 (2014).



comparative perspectives,<sup>18</sup> historical studies of IQ testing have usually focused on the nation as their main unit of analysis.<sup>19</sup> Such studies highlight the diversity of locally anchored actors and processes, indicating that the content of what is meant by “intelligence” and the techniques and scales for measuring it may vary by context. While the nation is also a key analytical unit in this volume, the national framework is supplemented and challenged by demonstrating that sub-national local units such as municipalities, or international units such as Scandinavia, have been important in shaping the history of IQ testing in Norway, as have networks that cut across local, national, and regional boundaries. By tracing the historical movement of tests, test-related knowledge, and people within and outside Norway, and by exploring the role IQ tests have played across various sectors in Norwegian society, this volume offers novel perspectives on the project of decentring canonized IQ test narratives. It is ultimately a book about IQ testing in Norway, rather than an attempt at narrating the Norwegian history of IQ testing.

The nation is still an important analytical unit in most of the chapters, owing to the fact that testing practices in large part have been connected to processes related to national culture and administration. For example, IQ tests have usually been explicitly translated and adapted for use on a Norwegian population and marketed to a Norwegian market. Test expertise and certification, likewise, have come to be organized on a national level. Most importantly, IQ testing has been assigned tasks

---

18 Carson, *Measure of Merit*; Mülberger ‘Biographies’; Mülberger, ed., ‘Mental Testing’; Bjørn Hamre, Tom Axelsson and Kari Ludvigsen, ‘Psychiatry in the sorting of schoolchildren in Scandinavia 1920–1950: IQ testing, child guidance clinics, and hospitalisation’, *Paedagogica Historica*, 55:3 (2019); Christian Ydesen, *The Rise of High-Stakes Educational Testing in Denmark 1920–1970* (Bern: Peter Lang Verlag, 2011), Christian Ydesen, Kari Ludvigsen and Christian Lundahl, ‘Creating an Educational Testing Profession in Norway, Sweden and Denmark, 1910–1960’, *European Educational Research Journal*, 12.1 (2013).

19 See also Paul Davis Chapman, *Schools As Sorters: Lewis M. Terman, Applied Psychology, and the Intelligence Testing Movement, 1890–1930* (New York: New York University Press, 1990); Hamilton Cravens, *Before Head Start: The Iowa Station & America’s Children* (Chapel Hill: University of North Carolina Press, 1993); Clyde Chitty, ‘Eugenics, Race and Intelligence’, in *Education* (London: Continuum, 2007); Jim Wynter Porter, *A “Precious Minority”: Constructing the “Gifted” and “Academically Talented” Student in the Era of Brown v. Board of Education and the National Defense Education Act* (Doctoral dissertation, Michigan State University, 2017); Laila Zenderland, *Measuring Minds: Henry Herbert Goddard and the Origins of American Intelligence Testing* (Cambridge: Cambridge University Press, 2001).

within nationally organized systems, such as the education system or the legal system.

From the outset, there are many reasons to consider Norway an interesting case to include in the international literature on the history of IQ testing. From early on, IQ testing has been connected to eugenics and used to identify “superior” and “inferior” individuals, races, and classes, often being criticized for contributing to the naturalization of social hierarchies. Norway is arguably of special interest in this regard. IQ testing was introduced and became institutionalized in the Norwegian school system, health services, and legal system during a period when Norway developed into what is often described as a Scandinavian-style social democratic welfare state, where the working class and the labor movement played leading roles. The rising welfare state was dominated by a strong ideal of social equality and integration, and a wide acceptance of state intervention in the socio-cultural sphere and the economy to achieve these social democratic ideals. This appears to be in sharp contrast to the social Darwinist ideas that IQ testing is—rightly or wrongly—often associated with.

Regarding IQ, “race”, and eugenics, it is clear that Norway, like the USA and other western countries, saw the rise of a eugenics movement as well as the racialization of its minorities in the early twentieth century. The role of race in Norwegian history is very different from the colonial powers of Europe and the USA, with their histories of slavery and racial segregation. The period when IQ testing was introduced into Norwegian society was strongly characterized by an outspoken assimilation policy towards minorities, which aimed to create an ethnically uniform nation.

As elsewhere, the rise of IQ testing in Norway is closely related to the development of universal schooling, and a desire to classify and sort children according to ability levels. Norway was, from the late nineteenth century, characterized by a particularly strong and long-lasting drive towards a unitary school system—a public school open to children of all social backgrounds, giving everyone the same access to higher education. This ideal was increasingly brought to life during the interwar years, and the introduction of IQ testing in the same period was intertwined with the implementation of the unitary school. This trend was reinforced in the post-war era by the Labor Party, the dominating political force which saw education politics as a key element in their

struggle to eradicate class distinctions and enhance social integration. IQ tests were mainly imported from the US, but in contrast, IQ testing in Norway became part of a social democratic welfare state. This makes Norway an interesting case as compared to the history of IQ testing in “core” countries, such as the colonial powers of Europe or the US, the latter with its free-market capitalism and strong meritocratic ideals combined with racial segregation and a weak labor movement.

Finally, and as shown in many of the contributions in this book, scale matters. IQ tests are commercial products, and the size of the Norwegian test market had significant implications for the development and marketing of IQ tests in Norway. A recurring topic in the history of IQ testing is the interaction between the development of IQ testing as both a commercial product and an “objective” scientific instrument, and the institutionalization and legitimization of psychology as a profession with a monopoly on the use of this instrument. In pre-war Norway, in contrast, the “ownership” of IQ testing was shared between psychiatrists, pedagogues, and psychologists. This development may perhaps be explained by the small scale of the Norwegian academic community, which is likely is also a contributing factor to why psychology was only institutionalized as a profession after World War Two in Norway.

## Existing Works on the History of IQ in Scandinavia

The last two decades have seen new studies dealing with the early twentieth-century history of intelligence and psychometrics in Denmark, Sweden, and Norway, including comparative and transnational perspectives.<sup>20</sup> Yet, in Norway this topic has so far been predominantly handled as a tangential aspect of other topics, such as the history of professions (psychiatry, psychology, pedagogy, or special needs education), education history, and the history of eugenics. These works focus on the period up to the 1950s within a primarily national and Scandinavian context.<sup>21</sup> In this book, in contrast, we have urged our

20 Axelsson, Rätt elev i rätt klass; Christian Lundahl, ‘Making testers out of teachers: The work of a Swedish State Research Institute 1946–1956’, *History of Education*, 48:5 (2019); Ydesen, ‘The rise’; Ydesen et al., ‘Differentiation’; Ydesen et al., ‘Creating’.

21 Froestad and Ravneberg, ‘Education Policy’; Per Haave, *Sterilisering av tatere 1934–1977. En historisk undersøkelse av lov og praksis* (Norges forskningsråd, 2020); Kari Ludvigsen and Åsmund Arup Seip, ‘The Establishing of Norwegian Child Psychiatry:

contributors to put IQ tests, testing practices, and related knowledge production at the centre of each case study, collectively beginning in the early twentieth century and continuing into the present. While the chapters will substantiate this project, we offer here an extended, but necessarily selective, summation of the chapters and their broad arguments. Alternatively, we invite you to delve straight into the individual contributions, all written as self-contained pieces.

## The Chapters of the Book

Chapter 1, by historian of science Ageliki Lefkaditou, begins by guiding us through the early implementation of IQ testing within education and psychiatry in the 1920s and 1930s. Lefkaditou shows how a group of progressive pedagogues, psychologists, and psychiatrists formed around the special school issue and the call for a Norwegian standardized IQ test. Their IQ-testing advocacy was characterized by ideals about care and control. Identification and segregation into special schools were said to enhance the quality of life and the education of the weakly gifted child, while at the same time relieving ordinary pupils and their teachers of a burden. The chapter culminates in an account of the activities of Norway's first school psychiatrist, Johan Lofthus, who translated, adapted, and normed a Norwegian version of the Stanford–Binet, known as *Osloprøven* [the Oslo test], recognized as the standard Norwegian intelligence test from 1931 to the 1950s.

In narrating the intricate story of early Norwegian efforts to adapt the Binet–Simon and Stanford–Binet intelligence scales, the chapter sheds important light on the test itself, introducing the practical and statistical acrobatics involved in defining a representative sample and establishing a national standard. Lofthus, for instance, made significant changes in the test content while maintaining its identity as a Stanford–Binet test through this process. The chapter also historically demonstrates how supposedly stable cut-off points between categories of children were

---

Ideas, Pioneers and Institutions', *History of Psychiatry*, 20.1 (2009); Bodil Ravneberg, *Normalitetsdiskurser og profesjonaliseringsprosesser: en studie av den spesialpedagogiske yrkesutviklingen 1880–1990* (Doctoral dissertation, University of Bergen, 1998); Eva Simonsen, *Vitenskap og profesjonskamp. Opplæring av døde og åndssvake i Norge 1881–1963* (Doctoral dissertation, UiO, 2000).

in flux and debated. Decisions about the transfer of a child to special education could be based on complex deliberations that, in addition to an IQ score, included the teacher's professional assessment, practical, institutional, and economic considerations, as well as the social standing and authority of the parent.

Chapter 2, by historians of education Christian Ydesen, Brit Marie Hovland, and Emma Vikström, maps out the closely knitted networks of Scandinavian actors and their mutual influence on intelligence testing in the interwar years. Centring on the Nordic School Association and their journal, we learn how influential psychiatrists, psychologists, and educators operated across Scandinavia and beyond, drawing inspiration and learning from each other in the early institutionalization of IQ testing in education. The chapter points to both differences and similarities in the special school system and in the use of IQ testing across Norway, Denmark, and Sweden in a time that saw the emergence of the foundations of unitary education. The idea of one-school-for-all that today is encapsulated in the Nordic model of education, as the authors point out, was dependent on the exclusion of those deemed to be uneducable. The authors show how IQ testing was discussed as one of the most effective solutions to sort children into special schooling, for perceived benefit both to the individual children and society at large.

Beyond this shared goal, however, there was no consensus about the role of IQ testing. The chapter identifies a line between a biologically grounded scientific and medical approach—that is, intelligence as hereditary and unchanging—and a pedagogical and educational approach that emphasized practical pedagogical knowledge and educational measures. But despite some disagreements, the interaction in the association's journal is characterized by mutual recognition and 'co-operation between the guardians of science and the teachers of the special school'.<sup>22</sup> IQ testing appears as the preferred means and medium of such cooperation, offering a standardized way of charting the potentiality of learners, and adding scientific objectivity to conversations between different professional judgments.

Chapter 3, by historian of science and medicine Svein Atle Skålevåg and legal scholar Linda Gröning, continues this theme of IQ tests as implicated within distinct fields of professional judgment, albeit within the confines of psychiatry and law. Specifically, they trace the

---

22 Elgström, cited in Ydesen, Hovland and Vikström, chapter 2, this volume.

role of intelligence testing in the history of criminal unaccountability through key Norwegian criminal law reforms from 1902 up until our own time. These reforms show not only changing conceptions of unaccountability, from the absence of reason to medical pathology, but also an increasing embeddedness of intelligence testing in assessments of criminal unaccountability. In fact, the authors propose that it was intelligence tests that brought conceptions of intelligence into law—first through the notion of mental age and, much later, from the 1970s onwards, expressed through IQ, and through later integration of ICD (International Classification of Diseases) classifications into legal definitions of unaccountability.

Psychiatrists were the first to employ intelligence tests in Norway. The Binet method, alongside various diagnostic practices, became a cornerstone of forensic-psychiatric assessments of intelligence and intellectual disability from the 1920s onward. The introduction of intelligence testing into forensic psychiatry and law are here shown to be tempered by existing disciplinary expertise and structures. We see how there is an ambivalence about the use of tests from both sides, and how the use of intelligence tests is followed by the insistence that the result must be understood in relation to a broader assessment. Nevertheless, intelligence tests pose as seemingly indispensable tools to aid difficult legal decision-making, imbuing a sense of objectivity into this process. With reference to current legal practice, for instance, Skålevåg and Grønning note that there are ‘cases where the guiding IQ limit still seems to be interpreted as an absolute demarcation, so that formalized IQ tests can determine who can be considered unaccountable and not.’

Chapter 4, by historian of medicine Per Haave and historian of science Jon Kyllingstad, traces how IQ testing and eugenics figured in the implementation of the Norwegian sterilization law from its inception in the 1930s up to the 1960s.

The 1934 Sterilization Act both legalized sterilization as a contraceptive method and paved the way for the involuntary sterilization of people with severe mental disabilities. Haave and Kyllingstad show how IQ tests, while not figuring in the letter of the law, functioned as a criterion in cases of the sterilization of people with a “mental deficiency” a historic precursor to the contemporary diagnosis of intellectual disability. The chapter presents the eugenic background of the act and lays out how IQ testing, together

with an assessment of what was called 'ethical level' and psychopathy, became important in psychiatric assessments of mental deficiency levels.

The deliberations involved in psychiatric discretionary assessments are highlighted through a court case that received much public attention in the mid-1950s. A woman who had been sterilized without her consent in 1942 went to court against the state and the psychiatrist who had diagnosed her as "imbecilic". She was later retested, and the contradictory higher new test results became a key bone of contention.

While the state and the psychiatrist were acquitted on the basis of "the progress of science", the case became a turning point in the practice of involuntary sterilizations, with the Sterilization Board becoming more restrictive. This was both due to the fear of new court cases, but also due to changed attitudes towards the rights of mentally disabled people and a declining support for eugenic sterilizations. Nevertheless, and at first sight paradoxically, it is concluded that this increased attention to the rights of people assumed to have a mental deficiency is likely to have led to an even greater emphasis on IQ scores in sterilization cases.

Chapter 5 by Isak Emberland, historian of professions, charts the role of psychotechnics and vocational testing in socially legitimizing psychological expertise, and cementing the psychological profession in Norway in the period 1925 to 1948, when a professional university exam in psychology was founded. Emberland offers important reminders of how IQ tests are part of other psychological tests and assessment practices. The Norwegian story is situated within the international development of applied psychology, and the chapter traces how this professional and popular movement contributed to the professionalization of what eventually became a clinical profession in Norway. Emberland shows how these international ideas were picked up and advocated by a diversity of Norwegian actors, including engineers. Psychotechnics was seen as a politically neutral scientific tool that could contribute to increased effectiveness, and help individuals find their place in society and work life. Psychotechnics, with its associated ideas, in turn created a market for psychological expertise.

A central point of analysis is the Psychotechnical Institute, part of the Oslo Vocational School, where we are also introduced to Helga Eng's work on the 'Goldsmith test' against a background of industrialization, and the emergence of the need for new forms of training beyond the

traditional master-apprentice relationship. The development of the test showcases how local needs inform test construction. We learn that intelligence testing played a role in psychotechnics, but only a supplementary one. While IQ scores were meant to predict school performance, psychotechnical testing explicitly probed for aptitudes relevant for specific crafts that ‘traditionally [were] suppressed or devalued in schooling’. This was related to a wider interwar critique of the school system as not sufficiently catering to the needs of the employment sector.

Chapter 6, by political scientist Kari Ludvigsen, examines the role of IQ testing in an urban Norwegian Educational Psychology office between 1953 and 1980 through a close analysis of yearly reports. During this period, the Norwegian school system went through a significant change. Reforms in the 1950s, characterized by interwar principles of segregated special education, led to a strong expansion of municipal special schools and classes for “slow learners”, as well as state schools for the disabled. This development reached a peak in the latter part of the 1960s. In parallel, the Educational Psychological Services developed into a nationwide service, and its most significant task from the start was the assessment of individual children for transfers to special education. In 1969 and 1976, there were school reforms that led to the segregation principle being replaced by a principle of integration—that all children should have access to education within the normal primary school based on their abilities.

The chapter sheds light on how such a testing practice was described, legitimized, and related to other tasks, priorities, and regulations. Early reports emphasized efforts to equip the office with sufficient testing instruments and expertise, as well as the undertaking of individual and group testing. An ever-increasing demand for individual assessments created a great deal of work pressure, and during the 1960s there was more emphasis on preventive measures and closer cooperation with schools and teachers. Reports from the late 1960s and early 1970s included some expressions of concern about the potentially stigmatizing effect of individual assessments. Throughout the period, IQ and personality testing were held up as important tools in difficult individual cases, but reports show greater emphasis on the role of the school and other environmental conditions in the child’s development.

Ludvigsen concludes, however, alongside other previous studies, that



the changes in the school system did not lead to any significant breaks in test use. Throughout the period, IQ tests were used pragmatically, along with other counselling and assessment tools, to meet growing demand from teachers and parents, and to solve changing tasks provided through national and local guidelines.

Chapter 7 by Håkon Caspersen, a social anthropologist, offers glimpses into the contemporary use of IQ tests, or rather *evnetesting*—tests of general ability, as they are frequently referred to—in the EPS today. While historically grounded, it builds in large part upon observational and interview-based material, and revolves around the use of the Wechsler Intelligence Scale for Children (WISC). This is a test that continues to act as an example of best practice in IQ testing internationally, and, since the late 1960s, it has been the most used IQ test in the EPS of Norway.

Throughout the chapter, we get a sense of how practitioners speak and learn to speak of, use, critically discuss, and understand these tests today. Here it is, perhaps surprisingly, not necessarily the full-scale IQ score that is of most interest. For some, test scores figure as an initial mapping tool in a wider assessment, while for others, these scores function as a last resort when other assessment procedures fail to elucidate the issue at hand. But regardless of frequency, the WISC remains a central tool of the trade, both a professional marker and a communicative device. Practitioners note the value of the test not only for its results—given perceived external expectations of testing from the schools or the psychiatric services—but also for the information gleaned in meeting with a child. The testing situation can become a time-efficient way to assess a child's difficulties firsthand even beyond the test.

Of special note is how in Norway, it is not only psychologists or psychiatrists who conduct IQ tests. Most employees in the EPS today come from a background in education or special needs pedagogy. The chapter points to regional differences, and the likely influence of disciplinary backgrounds on test practice—although the explicit question of whether the test is used differently by a psychologist or a special needs pedagogue is left unanswered.

Chapter 8, by the science and technology (STS) scholar Susanne Bauer, reveals the use of one of the Wechsler tests as part of a clinical intervention trial in a large interdisciplinary study on kindergarten

children. The study sought to assess the potential positive effect on cognitive abilities of eating fatty fish. It was funded and supported by the aquacultural industry, together with the Norwegian Research Council. The chapter shows how the test is pulled in as one element of a much larger set of techniques for quantification of a range of variables. The employment of a Wechsler test, Bauer shows, needs to be understood on the basis of the historical existence of this test system as a well-researched and validated tool, and, not least, one that has been standardized and translated into Norwegian. It becomes the best choice because it is readily available and understood as scientifically reliable, and hence easily plugged in to larger systems of knowledge production.

Bauer analyzes the function of tables and the assorted scientific boxing practices associated with statistics in this study, and how the Wechsler tests here join the “web of causation” at the core of statistical rationality and evidence-based reasoning. Importantly, the chapter notes the significance of the institutional placing of the research, and how educational institutions have long functioned as the “go-to field” for psychometrics and other research. This type of knowledge practice—intervention studies in educational settings—habituate public institutions to specific modes of evaluation, Bauer argues, that in sum enact an ableist project where enabling better performance has become part of pedagogy, care, and the public good. This is a dynamic that has been a feature of social policy and the welfare state from its beginning, and the employment of metrics such as IQ tests continues to lend credibility to such projects.

Chapter 9, by the historian of technology and business Gard Paulsen, takes us back to the post-World-War period and traces the travels, adaptation, translation, and use of the test that, in 1954, would be published as *The Norwegian Standard of Terman–Merrill’s Revision of the Binet Tests*. IQ tests are not just “scientific technology”; for most of their history, the chapter notes, they have also been commercial (and copyrighted) products, existing as books and record booklets that dictate procedural action. Here, it is the “shelf life” that is explicitly taken up, and in the Norwegian case, the longevity and popularity of a product that Norwegian publishers first considered unsellable. In fact, its very advertisement upon its publication emphasized restrictions on its sale—it was only medical doctors or other qualified personnel

that were meant to get their hands on the product. Furthermore, this was a product that had been partly standardized with the help of the work of students in pedagogy at the Pedagogical Research Institute, and that in its original American form was peddled to both teachers and psychologists alike.

Importantly, the chapter also brings us close to the materiality of the test, and the changes and logistical efforts—which even required the importing of a specific set of toys—necessary to complete this product and the work deemed necessary for it to deserve its title as “standard”. This process and its end product were nevertheless determined to be unsatisfactory by those involved. But once it was released, and in the context of a developing educational pedagogical service, Paulsen shows how the test quickly established itself as a standard test, not necessarily because it was especially good, but due to lack of other suitable alternatives.

Chapter 10, by historian of science Jon Kyllingstad, uncovers the conceptualizations of culture and ideas of representability at play in the Norwegian adaptation of WISC-R in the 1970s. The background of the Norwegian standardization was the EPS requirement for an instrument to measure general ability when assessing a child’s need for special education. “General ability” was here understood to serve as a measure of the child’s previous learning that, in turn, was based on both the child’s innate predispositions and their general life experiences. The chapter describes the criteria used for constructing a “representative” sample of pupils, on whom the Norwegian version of the test was tried out and normed (the chapters by Lefkaditou and Paulsen provide interesting comparative takes on earlier standardization processes of different tests). It shows how the adaptation process was based on the expectation that intelligence, as measured by IQ, would be lower among working-class children than among children of well-educated parents. However, while socio-economic status was a key criterion in the construction of the representative sample, cultural or linguistic differences related to ethnicity or language minority backgrounds were not taken into consideration.

This is conspicuous, because the original US version of WISC-R came into being against the backdrop of tense scientific and public debates about the cultural, social, and racial bias of IQ testing, and the Norwegian test developers had close relations with their American colleagues. Still, these controversial issues received no attention in the

Norwegian adaptation, and it seems to have been taken for granted that these debates had no relevance for Norway.

The chapter ends with a reference to studies published in the early 2000s of EPS assessment practices that highlight cases where minority-language children were tested with WISC-R, without taking into consideration the fact that they did not belong to the population that the test was adapted for. In several cases, this led to children with limited Norwegian language skills being misdiagnosed as having general language difficulties.

Chapter 11, by historian of professions Fredrik Thue, delves into a public and political controversy in the late 1950s regarding intelligence testing of schoolchildren in Oslo. The dispute was not about the role of IQ tests as a sorting mechanism for special education, but about the general use of group tests on first graders. While proponents emphasized the use of the tests as an instrument to aid the adaptation of teaching to the individual abilities of the child, critics called it an attack both on the child's personal integrity and parental rights.

Thue proposes that the IQ controversy should be seen in the context of a built-in tension in the Norwegian unitary educational system—one that, since its inception in 1884, has been seen as under democratic control from the municipality and parents, and at the same time as a tool for national integration and social equalization. Comprehensive Labor-Party-led reform work in the 1950s and 1960s led to parents' rights being overshadowed by macro-reform perspectives. In this light, the IQ controversy can be partly understood as a protest against centralization, technocracy, and Labor dominance. This may help explain why much of the criticism of intelligence testing in Norway came from the right and addressed the rights of the family and the individual vis-à-vis the state and scientific expertise. This contrasts with the US, where the criticism against IQ in the 1960s and 1970s often came from the left and focused on the racial and social biases of IQ testing.

Chapter 12, by the educational historian Kim Helsvig, explores the role of psychometric testing in the development of Norwegian school policy, comparing the rise of the social-democratic unitary school after World War Two with the OECD-inspired adaptation of the Norwegian school system at the turn of the millennium.

As the dominating political party of the 1950s, Labor sought to

prolong the unified mandatory public-school requirement from seven to nine years. The goal was both to capture the talent reserve in the population and to enhance national integration. The opposition feared a lowering of the overall level if all students were to attend the same secondary school. The last three years of school were therefore to be divided into tiered tracks. In 1954, the Institute of Educational Research at the University of Oslo was entrusted with the task of developing a test that could help advise a track in line with the students' abilities. Helsvig argues that this "maturity" test, and related psychometric expertise, became a decisive factor in legitimizing the reform. It soon lost significance, however, when the Parliament in 1963 abandoned the plan of a tracked system. During the next decade, school-policy-related psychometric research became strongly marginalized in Norway.

However, psychometrics made a strong comeback after 2001, when Norway's average score on the first Programme for International Student Assessment (PISA) survey indicated a Norwegian school system in crisis. This, combined with the rise of the ideal of an education system geared towards the production of competencies that could ensure the nation's competitiveness in the emerging knowledge economy, led to the revival of psychometry.

Like the maturity test of the 1950s, the PISA survey was meant to measure supposedly culturally neutral school-relevant skills. Both tests are very similar to IQ tests and were mainly developed as tools for political and bureaucratic control. Helsvig argues, however, that the maturity test was also developed with the pupils' interests in mind, to help them find an educational path in accordance with their abilities and interests. The psychometric wave half a century later was in contrast, mainly about governance and the aim of enhancing the population's future skills, as well as the economic competitiveness of the nation.

Chapter 13, by historian of science Jim Porter, takes Norway as a vantage point for a comparative view on the history of IQ testing in US schools and, based on insights from the chapters of this volume, discusses similarities and differences between the nations.

Around the start of the twentieth century, both nations saw an expansion of public schooling, along with increasing demands for more efficient and science-based schools, and the rise of segregated special education for students who had difficulty following the normal progression in school. Progressive educational reformers in both

countries embraced IQ tests as scientific instruments to help develop an education more adapted to the abilities of the individual child, and to single out children for special education, thus defining the boundary of normality. IQ testing in the Norwegian school system continued throughout the century to be mainly linked to special education, and only from the 1950s onwards on a nationwide scale.

In the US, however, IQ testing expanded much more rapidly and attained far more extensive functions. Porter connects this US breakthrough to three interconnected factors: 1) Racist and classist beliefs about the unequal educability of different social groups, which led to demands for differentiation; 2) educational policies which, in turn, led to organizational differentiation in schools (“tracking”), and thereby stimulated a demand for testing; and 3) market forces which affected educational policy and the rise of a testing business. The chapter shows how the mutually reinforcing relationships between these factors developed over time in the US and helped increase the significance of IQ testing in US schools, from the 1920s and into the decades after World War Two. It also discusses its difference from the Norwegian trajectory due to basic societal, economic, and political factors. Norway was too small to become a profitable market for a testing industry. The societal relevance of “race” and class were very different in Norway, with a less heterogeneous population and a strongly assimilationist policy towards minorities, coupled with a social democratic version of capitalism, and a unitary public school as part of a welfare state that developed very differently from the US.

Chapter 14 continues this comparative effort, standing as a companion piece to this introduction and the individual chapters by further elucidating a few overarching points regarding the history of IQ testing in Norway, partially glimpsed throughout this book. These include the comparatively limited breakthrough of testing in Norway versus the US, and the relative absence of discussions of “race” in the Norwegian history of IQ testing. Here, we also consider the oft-noted theme of the pragmatic and practical use of these tests, and their role in the reification of intelligence, as seen from Norway but with similarities to other countries.

The volume concludes with an Afterword by Annette Mülberger, a historian of psychology.

## A Multidisciplinary Approach

Given the historically rooted but multidisciplinary nature of the book, the reader will notice different, even divergent, analytical takes on IQ testing, their scientific legitimacy, and social value. Despite some coordination, we have approached IQ testing in Norway from different starting points that together provide a multifaceted portrayal of the topic. In addition to being a contribution to international literature on the history of IQ testing, the individual chapters can be read as contributions to several other specialized and interdisciplinary discussions—including the role of IQ testing in criminal law today, the medicalization of schooling, as well as more general discussions on testing and assessment procedures in education and school psychology. Other themes that weave through the volume are the relationship between commodification and science, the legitimization of professional knowledge and power, social aspects of statistical categorization, changing conceptions of care, bureaucratization, the welfare state, and technological implementation. In spite of this far-ranging approach, there are still many gaps left to be filled in the mapping of this history, such as post-war testing in the military, the use of such tests in work and business, and, not least, the different usages of IQ tests in the healthcare system, in population-level statistics, or in scientific knowledge production more broadly (although Bauer, in chapter 8, explores the latter to some extent).

## A Note on Historical Terms

The reader will come across a range of historical terms that today might be considered offensive, or at least problematic, due to their negative social connotations. They were used as descriptive markers of individuals or groups of people for categorization; examples include “backward”, “idiot”, “imbecile”, “debile”, “mentally retarded”, “laggards”, “mentally deficient”, and “handicapped”, among others. Their Norwegian equivalents are left out above, as are usages that get lost in translation. The word *rase* [race], for instance, can silence a seminar room of first-year Norwegian anthropology students (‘Can we even use that term?’), while prompting politically charged discussions

amongst their counterparts in Britain or the US.<sup>23</sup> Even the meaning of terms that seemingly remain consistent over time, such as “IQ”, change and develop, as seen above. While one speaks of IQ scores in the 1920s and the 2020s, they do not refer to the same thing across time. It is in recognition of such differences and possible contentions that the contributors have taken pains to use the terms employed historically, to ensure context and accuracy of meaning in line with changes over time. Such terms are indicated by double quotation marks upon first use in each chapter, and the original Norwegian term is often included alongside its English translation.

## Bibliography

- American Psychiatric Association, *Diagnostic and Statistical Manual: Mental Disorders* (Washington DC: Mental Hospital Service, 1952).
- Andreasen, Thor Herman and Øyvind Fallmyr, ‘Kan vi leve med manglende norsk normering på WISC, WAIS og WPPSI?’ (Psykologitidsskriftet.no, 2006).
- Axelsson, Tom, *Rätt elev i rätt klass. Skola, begåvning och styrning 1910–1950* (Doctoral dissertation, Linköpings Universitet, 2007).
- Axelsson, Tom, *Att konstruera begåvning—Debatten om IQ* (Malmö University electronic publishing, 2012). <https://doi.org/10.24834/educare.2012.1.1189>
- Bergstrom-Katz, Sasha, *On Intelligence Tests: Psychological Objects and Their Subjects* (PhD thesis, Birkbeck University of London, 2023).
- Boake, Corwin, ‘From the Binet–Simon to the Wechsler–Bellevue: Tracing the History of Intelligence Testing’, *Journal of Clinical and Experimental Neuropsychology*, 24.3 (2002), 383–405, <https://doi.org/10.1076/jcen.24.3.383.981>
- Byford, Andy, ‘The mental test as a boundary object in early-20th-century Russian child science’, *History of the Human Sciences*, 27.4 (2014), 22–58, <https://doi.org/10.1177/0952695114527598>
- Carson, John, ‘Mental Testing in the Early Twentieth Century: Internationalizing the Mental Testing Story’, *History of Psychology*, 17.3 (2014), 249–255, <https://doi.org/10.1037/a0037475>

---

23 For a detailing of the history of race as a scientific category and its specificity of use in Norway, see Jon Kyllingstad, *Rase: en vitenskapshistorie* (Oslo: Cappelen, 2023).



- Carson, John, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750–1940* (Princeton: Princeton University Press, 2007).
- Chapman, Paul Davis, *Schools As Sorters: Lewis M. Terman, Applied Psychology, and the Intelligence Testing Movement, 1890–1930* (New York: New York University Press, 1990).
- Chitty, Clyde, *Eugenics, Race and Intelligence in Education* (London: Continuum, 2007).
- Cravens, Hamilton, *Before Head Start: The Iowa Station & America's Children* (Chapel Hill: University of North Carolina Press, 1993).
- Danziger, Kurt, *Constructing the Subject: Historical Origins of Psychological Research* (Cambridge: Cambridge University Press, 1990).
- Daston, Lorraine, 'Introduction: The Coming into Being of Scientific Objects', in *Biographies of Scientific Objects*, ed. by Lorraine Daston (Chicago: University of Chicago Press, 2000).
- Evans, Bonnie, *The Metamorphosis of Autism: A History of Child Development in Britain* (Manchester: Manchester University Press, 2017).
- Froestad, Jan and Bodil Ravneberg, 'Education Policy, the Norwegian Unitary School and the Social Construction of Disability', *Scandinavian Journal of History*, 31.2 (2006), 119–143.
- Gould, Stephen J., *Mismeasure of Man*, (New York: W. W. Norton & Co., 2<sup>nd</sup> ed., 1996).
- Haave, Per, *Sterilisering av tater 1934–1977. En historisk undersøkelse av lov og praksis* (Norges forskningsråd, 2000).
- Hanson, F. Allan, *Testing, Testing: Social Consequences of an Examined Life* (Berkeley: University of California Press, 1993).
- Hamre, Bjørn, Tom Axelsson and Kari Ludvigsen, 'Psychiatry in the sorting of schoolchildren in Scandinavia 1920–1950: IQ testing, child guidance clinics, and hospitalisation', *Paedagogica Historica*, 55.3 (2019), 391–415, <https://doi.org/10.1080/00309230.2019.1597137>
- Herrnstein, Richard J. and Charles Murray, *The Bell Curve: Intelligence and Class Structure in American Life* (New York: Free Press, 1994).
- Kyllingstad, Jon Røyne, *Rase: en vitenskapshistorie* (Oslo: Cappelen, 2023).
- Leopoldoff, Irina, 'A psychology for pedagogy: Intelligence testing in USSR in the 1920s', *History of Psychology*, 17.3 (2014), 187–205, <https://doi.org/10.1037/a0035954>
- Ludvigsen, Kari and Åsmund Arup Seip, 'The Establishing of Norwegian Child Psychiatry: Ideas, Pioneers and Institutions', *History of Psychiatry*, 20.1. (2009), 5–26, <https://doi.org/10.1177/0957154X08089847>

- Lundahl, C., 'Making testers out of teachers: The work of a Swedish State Research Institute 1946–1956', *History of Education*, 48.5 (2019), 646–663, <https://doi.org/10.1080/0046760X.2019.1565422>
- Mülberger, Annette, 'Biographies of a Scientific Subject: The Intelligence Test' in *Oxford Research Encyclopedias: Psychology* (Oxford: Oxford University Press, 2020), <https://doi.org/10.1093/acrefore/9780190236557.013.694>
- Mülberger, Annette, 'The need for contextual approaches to the history of mental testing', *History of Psychology*, 17.3 (2014), 177–186, <https://psycnet.apa.org/doi/10.1037/a0037487>
- Mülberger, Annette, ed., 'Mental Testing after 1905: Uses in Different Local Contexts' (Special Issue), *History of Psychology*, 17.3 (2014).
- Porter, Jim Wynter, A "Precious Minority": Constructing the "Gifted" and "Academically Talented" Student in the Era of Brown v. Board of Education and the National Defense Education Act (Doctoral dissertation, Michigan State University, 2017).
- Ravneberg, Bodil, *Normalitetsdiskurser og profesjonaliseringsprosesser: en studie av den spesialpedagogiske yrkesutviklingen 1880–1990* (Doctoral dissertation, University of Bergen, 1998).
- Rose, Nikolas, *Governing the Soul: The Shaping of the Private Self* (London: Free Association Books, 2<sup>nd</sup> ed. 1999).
- Sattler, Jerome M., *Assessment of Children: Cognitive Foundations and Applications*, 6<sup>th</sup> revision (Sattler Publishing Inc., 2018).
- Simonsen, Eva, *Vitenskap og profesjonskamp. Opplæring av døde og åndssvake i Norge 1881–1963* (Doctoral dissertation, UiO, 2000).
- Strand, Nina, 'WISC-III skaper uro' (Psykologitidskriftet.no, 2025).
- Sutherland, Gillian, *Ability, Merit and Measurement: Mental Testing and English Education 1880–1940* (Oxford: Clarendon, 1983).
- Wasserman, John D., 'A History of Intelligence Assessment: The Unfinished Tapestry' in *Contemporary Intellectual Assessment: Theories, Tests, Issues*, ed. by Dawn P. Flanagan and Erin M. McDonough (Guilford, 2022).
- Wechsler, David, *The Measurement of Adult Intelligence* (Baltimore: Williams & Wilkins, 1939).
- Wooldridge, Adrian, *Measuring the Mind: Education and Psychology in England, c. 1860–1990* (Cambridge: Cambridge University Press, 1994).
- Ydesen, Christian, *The Rise of High-Stakes Educational Testing in Denmark 1920–1970* (Bern: Peter Lang Verlag, 2011).
- Ydesen, Christian, Bjørn Hamre and Karen Egedal Andreasen, 'Differentiation of Students in the Early Danish Welfare State: Professional Entanglements Between Educational Psychologists and Psychiatrists', *Nordic Journal of Educational History*, 5.1 (2018), 73–96.

- Ydesen, Christian, Kari Ludvigsen and Christian Lundahl, 'Creating an Educational Testing Profession in Norway, Sweden and Denmark, 1910–1960', *European Educational Research Journal*, 12.1 (2013), 120–138, <https://doi.org/10.2304/eerj.2013.12.1.120>
- Young, Jacy L., 'Test or Toy? Materiality and the Measurement of Infant Intelligence', *History of Psychology*, 18.2 (2015), 103–118, <https://doi.org/10.1037/a0038862>
- Zenderland, Laila, *Measuring Minds: Henry Herbert Goddard and the Origins of American Intelligence Testing* (Cambridge: Cambridge University Press, 2001).
- Øyhovden, Asbjørn, Connie Bentzrud, and Olav T. Hustad Wold, 'Metoderapport SKUP 2017: IQ Testing og Diagnosehelvete' in *SKUP: Stiftelsen for en kritisk undersøkende presse* (2017).
- Øyhovden, Asbjørn and Connie Bentzrud, 'Slo alarm om IQ: Barn og unge kan ha blitt stemplet som utviklingshemmede etter feil' (TV2.no, 2017).
- Øyhovden, Asbjørn and Connie Bentzrud, 'Helseministeren vil rydde opp i IQ-testing' (TV2.no, 2017).
- Øyhovden, Asbjørn, 'Strid om ansvaret for at barn og unge kan ha fått feil IQ-diagnose' (TV2.no 2017).

# 1. Between Change and Stability: The Thorny Adaptation Process of the Stanford–Binet Intelligence Scale in Norway

*Ageliki Lefkaditou*

---

‘The school with the beautiful name, which embraces *all* people’s children, the poor and the rich, should be more than hitherto also for the *weak* as well as the strong. Only then does it rightfully bear its name: *Folkeskolen* [People’s school—ordinary public primary school], *the school for all people*’,<sup>1</sup> wrote the teacher Julie Monrad Jacobsen, a pioneer of special education in Norway.<sup>2</sup> This was the concluding statement of her presentation on the *særskolen* [special school] at the Eleventh Nordic School Meeting in Oslo in August 1920. For Monrad Jacobsen, differentiation of the *svakt begavede* [weakly gifted] from the ordinary seven-year public school for all children aged seven to fourteen years old was seen as a precondition for making these children ‘content

---

1 Julie Monrad Jacobsen, ‘Særskolen, Hvad Den Er, Hvad Den Vil, Og Hvordeles Den Arbejder’, in *Elleve Nordiske Skolemøte i Kristiania 5–8 August 1920*, ed. by O. J. Hoversholm, Magnus Alfsen and G. E. Mæhlum (Kristiania: Fabritius, 1921), pp. 92–106 (p. 106). In this chapter, I use the term “people’s school” as a translation of the Norwegian *Folkeskolen* where the emphasis is on the political ideal of a school for all social classes. When it comes to the stage of education, I use the term “primary school”, although the Norwegian school included the first class of lower secondary, while the term “public” refers to the political decision of state-supported foundational education.

2 Ruth Frøyland Nielsen, Jacobsen’s successor at Ullevålveien’s special school, described her as the Grande Dame of special schools in Norway and internationally. Ruth Frøyland Nielsen, ‘Julie Monrad Jacobsen Død’, *Vor Skole*, 49.1–51/52 (1963), 22–23. See also, Gro Hagemann, *Skolefolk: Lærernes Historie i Norge* (Oslo: Ad Notam Gyldendal, 1992), p. 178.

with themselves' and 'the society content with them'.<sup>3</sup> She ended, though, by talking about a matter that deserved special treatment—how to determine who should be transferred to special schools. She informed her colleagues that the transfer process varied among countries and municipalities, but among the various schemes used, 'the most hypermodern were based on Binet–Simon's intelligence scale'.<sup>4</sup> Monrad Jacobsen lamented the lack of a specially adapted version for local circumstances, and called for closer collaboration with doctors, emphasizing that everyone would benefit from using such trustworthy tools.

By the time of Monrad Jacobsen's call, intelligence testing of schoolchildren with the aim of differentiating students considered of normal mental development from those deemed abnormal was international practice.<sup>5</sup> The scale that psychologist Alfred Binet and his research assistant, physician Théodore Simon, first published in France in 1905 heralded the era in which such tools became the sought-after scientific answer to ever-expanding educational needs.<sup>6</sup> Already

---

3 Jacobsen, 'Særskolen', p. 97. In this chapter, the language referring to persons with disabilities, and especially intellectual disabilities, is as close as possible to the language used historically, but even with this understanding it remains difficult and hurtful. This choice reflects my intention not to make the historical context easier to accept, but instead to reveal the derogatory connotations and discriminatory practices.

4 Jacobsen, 'Særskolen', p. 103.

5 For the history of intelligence testing from the perspective of several national and cultural contexts, see *Mental Testing after 1905: Uses in Different Local Contexts*, ed. by Annette Mülberger, special issue of *History of Psychology*, 17.3 (2014).

6 In the years to follow the publication of the Binet–Simon scale, the scale did not remain static. From an unordered set of 30 items aiming to assess fundamental mental functions such as language, memory, and imagination, Binet and Simon revised it to a scale organized by age and increasing difficulty in 1908, with the number of items reaching 54 in its final 1911 version. The scale then traveled to the US with Henry Goddard, psychologist, eugenicist, and director of the Vineland Training School for feeble-minded boys and girls in New Jersey, and was standardized by the Stanford psychologist and eugenicist, Lewis Terman, in 1916, to become the Stanford–Binet scale (or test).

Newer studies show how despite the test's spread, critique to its practicality, ability to capture mental functioning and cultural dependency was mounting. For an overview of the development of intelligence testing as a scientific subject, see Annette Mülberger, 'Biographies of a Scientific Subject: The Intelligence Test', in *Oxford Research Encyclopedias: Psychology* (Oxford: Oxford University Press, 2020). Seminal works on the history of North American and North American versus French testing practices come respectively from Leila Zenderland, *Measuring Minds: Henry Herbert Goddard and the Origins of American Intelligence Testing* (Cambridge, UK: Cambridge, 1998) and

in the first two decades following the scale's publication, processes of adaptation—including translation, reordering of items, development of new items, and production of test materials—were underway in several countries.<sup>7</sup> These processes, often leading to a version of the scale described as the local standard, ensured that it remained both stable enough to retain its scientific veneer, and malleable enough to respond to local conditions.

This chapter examines the local efforts to develop a Norwegian version of the original Binet–Simon scale and its most popular revision by Lewis Terman, known as the Stanford–Binet. It starts with the earliest translations and uses of the scale in 1913, and concludes in 1931 with the publication of what contemporaries acknowledged as the first local standard by psychiatrist Johan Lofthus.<sup>8</sup> While historical studies in other disciplines have shown how standardization became the site for negotiating the internal workings of science alongside national and international politics, such inquiries are still lacking in the history of psychology.<sup>9</sup> Drawing on hitherto unexamined archival material and a close reading of the scale's adaptations, this chapter argues for a broader

---

John Carson, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750–1940* (Princeton: Princeton University Press, 2007).

- 7 Despite the widespread use of intelligence testing, Mülberger describes how it was surrounded by controversies from the very beginning culminating in what historians have called “IQ test wars” in the 1970s; see Mülberger, ‘Biographies’, pp. 1–2. In her detailed examination of intelligence test kits as archives, carrying clues of their multiple transformations, Sasha Bergstrom-Katz presents a rich history of alternatives to the Binet–Simon scale, and explains how this was in continuous development; see ‘On Intelligence Tests: Psychological Objects and Their Subjects’ (Unpublished doctoral thesis, Birkbeck, University of London, 2023), pp. 66–107.
- 8 The first to write a detailed account on the work of Johan Lofthus in relation to intelligence testing and its connection to eugenic ideas was Håkon Havik, *Mål Og Mening: Idéhistorisk Blikk På Johan Lofthus’ Intelligensmålinger i Norge 1918–1939* (Oslo: Universitetet i Oslo, 2006).
- 9 The work of Iris Clever in the history of physical anthropology showcases the importance, even if often rhetorically employed, of standardizing measurement methods for unifying this research field, and shows how this remained an elusive ideal caught between the forces of nationalism and internationalism. See, ‘Miriam Tildesley and the Anthropological Politics of Standardizing Racial Measurements’, *Perspectives on Science* 30.1 (2022), 13–47. Coreen McGuire writes on how the introduction of measurement techniques, including IQ in schools, in interwar Britain was an attempt to standardize the boundaries of disability, even if such standardization faced difficulties and was resisted; see *Measuring Difference, Numbering Normal: Setting the Standards for Disability in the Interwar Period* (Manchester: Manchester University Press, 2020).

understanding of the processes of standardization beyond statistical adjustments, translations, and uniform procedures.<sup>10</sup> It thus reveals how local adaptations depended on crossing geographical boundaries, securing financing, forming professional alliances, negotiating decision-making power, and intervening in the public debate with politically timely questions and pragmatic solutions. For some proponents of the scale, Norway's progressive educational system would find a reliable solution to school overcrowding, providing education adapted to individual needs, and sorting out the weaker students (i.e., removing them from the mainstream system).<sup>11</sup> For others, with the scale's authority residing in countries that were presumed to be scientifically more advanced, the adaptation would infuse the Norwegian scientific scene with experimental and empirical vigor.<sup>12</sup> In this sense, this chapter shows how test standardization was part and parcel of a complicated but not conflicting relationship between care and control—for science, society, and individuals.

## A School for All, with Exceptions

The history of various kinds of special education institutions in Norway for the children considered to be deviating from normality, including normal intellectual ability, goes back to the mid-nineteenth century. As elsewhere, these were initially private philanthropic establishments inspired by Enlightenment ideals. However, a series of reforms and new laws sought to regulate this landscape during the final two decades of

---

10 In this chapter, the terms “standardization” and “standard” are historical terms used by contemporaries and differ from current conceptions of standardization, especially in the elementary statistical methods employed. On standardization and adaptation, see also Paulsen, chapter 9, this volume.

11 Historical literature has previously discussed how the inclusive Norwegian school system—the first in Europe to move towards a common foundation education open to all regardless of class—depended on the exclusion of the abnormal or deviant children, often with the use of intelligence testing. See for example, Jan Froestad and Bodil Ravneberg, ‘Education Policy, the Norwegian Unitary School and the Social Construction of Disability’, *Scandinavian Journal of History*, 31.2 (2006), 119–43.

12 As we will see later in the chapter, among the proponents of the scale were the ones who advocated for the establishment of psychology as an independent discipline with an emphasis on experimental psychology; see Svein Magnussen, Bjørn Killingmo and Siri Erika Gullestad, *Klinikk og Laboratorium: Psykologi i Hundre År* (Oslo: Universitetsforlag, 2009), pp. 17–43.

the century. From 1881, state-owned institutions called *åndssvakeskoler*—roughly equivalent to the training schools for “feeble-minded” children in the USA—were responsible for the teaching of the educable *åndssvake* [feeble-minded].<sup>13</sup> The law expressed optimism about the benefits of education, but those considered ineducable were placed in separate institutions, either care homes or workhouses, and in some cases were simply sent back to their homes.<sup>14</sup> Reform schools were established in 1896 for the children and youth seen as neglected and/or exhibiting asocial behavior, with two broad categories including those characterized as psychopaths and feeble-minded.<sup>15</sup> The legal foundation for the establishment of the special schools that Monrad Jacobsen worked at were the *Lov om Folkeskolen i Kjøbstæderne* [Urban People’s School Act] and *Lov om Folkeskolen paa Landet* [Rural People’s School Act] of 1889.<sup>16</sup> The differentiation between these two almost identical

13 The law that established schools for the educable deaf, blind, and feeble-minded children is known as *Lov om abnorme Børns Undervisning* [The Abnormal School Act] of 1881. The foundational work on the history of special education in Norway, especially in relation to the state-owned training schools for the feeble-minded, is by Eva Simonsen, *Vitenskap og Profesjonskamp: Opplæring av Døve og Åndssvake i Norge 1881–1963*, 2 (Oslo: Unipub forlag, 2000). In this chapter, the terms “feeble-minded/feeble-mindedness” and “mentally defective/mental deficiency” are used interchangeably to interpret the complex historical term *åndssvake*. When I refer to current understanding, I use the term “intellectual disability”.

14 On the persistent differentiation of the educable from the ineducable pupils in Norway up until the 1970s, see Gøril Moljord and Hilde Bondevik, ‘Constructions of the In-/Educable: A Nordic Outlook on Changing Legal and Educational Statuses of Pupils with Disabilities’, in *The Nordic Education Model in Context: Historical Developments and Current Renegotiations in Studies in Curriculum Theory*, ed. by Daniel Tröhler (New York: Routledge, 2023), pp. 56–74.

15 The role of these institutions was to protect society from the criminal inclinations of young boys/men, and the girls’/women’s lax morals; see Per Haave, ‘Psykiatrien og De Vanskeligste Skolehjemsbarna: Tanke og Handling Mellom To Kriser’ (forthcoming). In his work on the plans to set up island institutions for ‘degenerated feeble-minded’ men and women in Sweden, Thomas Barow confirms these gender-based stereotypes with men described as ‘potentially criminal’ and women as ‘unreliable in terms of sexual behavior’; see “‘Excluded from the Outside World’: Planning and Discarding Island Institutions for “Degenerated Feeble-Minded (Asocial Imbeciles)” in 1920s Sweden’, in *Islands of Extreme Exclusion: Studies on Global Practices of Isolation, Punishment, and Education of the Unwanted*, ed. by Bjørn Hamre and Lisa Villadsen (Leiden: Brill, 2024), pp. 107–09 (p. 112). The gendered dimension of these institutions is further evident in their almost always admitting only inmates of one gender.

16 *Lov om Folkeskolen for Kjøbstæderne: Stockholms Slot den 26de Juni 1889* (Kristiania: P. T. Mallings Boghandels Forlag, 1889); *Lov om Folkeskolen paa Landet: Stockholms Slot den 26de Juni 1889* (Kristiania: P. T. Mallings Boghandels Forlag, 1889). While the laws



laws mirrored the large disparities between towns and the countryside, including the dispersed population and the prevalence of the New Norwegian language in rural areas.<sup>17</sup> Nevertheless, both laws included the provision to organize separate instruction for children whom the school authorities did not find to be fit for the ordinary school.<sup>18</sup> Special schools thus remained part of the ordinary primary school and were under municipal control. However, they were mostly a phenomenon of the big towns,<sup>19</sup> with children from the countryside in need of special education ending up at the state-owned training schools for the feeble-minded.<sup>20</sup>

These school reforms were the result of a long process of democratization, economic change and secularization in Norwegian society. They enjoyed broad political support, though not without controversy, and reflected a strong national sentiment, aspirations to support social mobility, and a turn towards more practical education. When it came to the sorting of schoolchildren, however, contemporary actors acknowledged that there were numerous borderline cases, both

---

restricted state support to people's schools, private schools continued to operate, although their numbers were in decline.

- 17 On the developments that led to the two versions of the law and the different ways in which schools operated between towns and countryside, such as big institutions with age-based classes and twice as many hours of instruction in the former, see one of the few overviews of Norwegian school history in English by Val D. Rust, *The Democratic Tradition and the Evolution of Schooling in Norway* (New York: Greenwood Press, 1989), pp. 106–17. In addition to the urban/rural distinction, the laws featured a gendered dimension with boys receiving more education in subjects such as mathematics and girls in home economics. For further discussion of this and other differences in the curriculum, as well as on the effects of progressive education reaching Norway, see Harald Thuen, *Den Norske Skolen: Utdanningssystemets Historie* (Oslo: Abstrakt forlag, 2017), pp. 88–127.
- 18 *Lov om Folkeskolen for Kjøbstæderne*, §6, p. 2 and *Lov om Folkeskolen paa Landet*, §7, p. 5.
- 19 A statistical survey of schoolchildren who attended special schools, or classes, in the period 1919–1923 in Norway mentions the cities and towns of Oslo, Trondheim, Bergen, Drammen and Arendal, although the phenomenon somewhat expanded in the subsequent years; see Eilif Gjermoe, 'Særskolestatistikk', *Hjælpskolan Værneskolen Særskolen* 6.1 (1927), 1–43.
- 20 Simonsen writes that the director for the state-owned training schools for the feeble-minded from 1913 to 1928, Ludvig Kragtorp, was actively trying to reduce expenses and thus preferred to have empty places in these institutions rather than to fill them with children whom he considered to have little potential. He further put an end to the transfer of weaker students who did not progress at the municipal special schools to the state schools, and considered the latter as catering to the same kind of children, only with state funding. This resulted in blurring even further the boundaries between the two institutions. *Vitenskap og Profesjonskamp*, pp. 213–24.

when they sought to move children from the ordinary schools and when they assigned pupils to the different kinds of special education institutions.<sup>21</sup> The overlap between those characterized as weakly gifted and the many degrees of mental disability or “difficult” behavior made the decisions complex and controversial. The ongoing debate revolved around whether it was scientifically or socially defensible to construct a separate—for some even privileged—category of children with intellectual disability attending special schools.<sup>22</sup> Others, especially ordinary school teachers, viewed this move as unnecessarily stigmatizing, and politically at odds with creating a common school for all.<sup>23</sup>

All objections notwithstanding, a group of teachers, doctors and psychologists formed around the cause for segregated special schools, and supported Monrad Jacobsen’s portrayal of intelligence testing as a new, promising tool in separating schoolchildren.<sup>24</sup> By the end of the Eleventh Nordic School Meeting in 1920, the *Nordisk særskoleforbund* [Scandinavian Special Schools Union] was established with representatives from Sweden, Denmark, and Norway. The aim was to ‘bring together individuals interested in education in the Nordic countries to collaborate in support of the weakly gifted, primarily through schools (classes) and/or associated social work’.<sup>25</sup> Monrad Jacobsen headed

---

21 On the debates around borderline cases, and how the sorting out of schoolchildren intersected with medical and pedagogical discourses, pessimism about the outcomes of teaching at the training schools for the feeble-minded, and international pedagogical reforms, see Bodil Ravneberg, ‘Grensebarna i Skolen. Utskillingen av “Sinkene” fra Normal- og Åndssvakekategoriene Tidlig på 1900-tallet’, in *Funksjonshemming, Politikk og Samfunn*, ed. by Jan Froestad, Per Solvang and Mårten Söder (Oslo: Gyldendal akademisk, 2000), pp. 112–33.

22 Asbjørn Birkemo mentions a few cases of resistance to intelligence testing, but a more systematic investigation is required; see *Pedagogisk-Psykologisk Rådgiving* (Oslo: Universitet i Oslo, 2009), pp. 21–49.

23 Froestad and Ravneberg emphasize how little support the municipal special schools received from the state, which treated them as ‘an unfortunate undermining of the idea of a unitary, homogeneous primary school for all (normal) children’ and by ordinary school teachers ‘who were traditionally sceptical [skeptical] about the idea of increased differentiation between the students’, ‘Education Policy’, p. 130. They add that teachers at the state special schools were equally worried about losing their status and role because of these new schools, and thus the schools relied on alliances between special school teachers, parents, and doctors.

24 In this chapter, I use the terms “segregation” and “segregated” to underscore the consequences that this practice of “sorting out” might have had on young children, although I have not been able to locate historical sources that include their voices.

25 ‘Nordisk Særskoleforbund’, in *Oslo Særskole: Gjennom 50 År*, ed. by Signe Gran and

both the union and its Norwegian division.<sup>26</sup> In November 1920, the *Kristiania Særskoleforening* [Kristiania, henceforth Oslo, Special Schools Association] was established with a steering board led by the first professor of psychiatry, Ragnar Vogt.<sup>27</sup> Other board members included Monrad Jacobsen, psychologist Martin Luther Reymert, consultant to the Ministry of Church and Education Otto Grennes, and special school teacher Anna-Margrethe Hansen. Seventy more members joined this association, which was at the time responsible for supporting the interests of two special schools with their own buildings.<sup>28</sup> These were Ullevålveien's school, placed in the more affluent western part of Oslo, and Sarsgaten's school in the working-class-dominated eastern part, with about 500 students and 39 teachers in total.

The Oslo Special Schools Association quickly sent an application to *Det Videnskabelige Forskningsfondet av 1919* [The Scientific Research Fund of 1919] seeking support for 'the national task [...] to standardize intelligence tests for Norwegian children'.<sup>29</sup> Since the first introduction of the Binet-Simon scale in 1905, such tests 'now constitute (in connection with further special tests) the most important objective standard for states and municipalities in terms of identifying and placing defective children and individuals'.<sup>30</sup> They argued that 'a swift resolution of this task is of the utmost practical significance for Norwegian homes and schools [...] which lacked exact means for the determination of children and youth's "general intelligence" in relation to a scale graded according to "intelligence age" and standardized for Norwegian conditions'.<sup>31</sup>

Following this, the widely circulated conservative newspaper *Aftenposten* presented the lack of standardized tests as a sign of Norwegian backwardness when it came to advances in experimental psychology.<sup>32</sup> Ragnar Vogt, interviewed in the same article, referred

---

Julie Monrad Jacobsen (Oslo: Thronsen & co, 1943), pp. 108–10 (p. 109).

26 'Beretning om Nordisk Særskoleforbunds virksomhet i Norge', *Skolebladet* 24.36 (1921), 532–33.

27 *Ibid.*, p. 532.

28 Gran and Monrad Jacobsen, *Oslo Særskole*, pp. 17–24.

29 The Scientific Research Fund was a free, independent fund established by the Norwegian parliament in 1919 with the aim to promote scientific work of all kinds. The Association's application was published in full in the teacher's magazine, 'Til Styret for Forskningsfondet av 1919', *Skolebladet* 24.44 (1921), 663–64.

30 *Ibid.*, 663.

31 *Ibid.*.

32 'Standardisering af Intelligensprøver for Norske Barn', *Aftenposten*, 543, 12 November 1921, p. 2.

again to standardization as a national task for both Norwegian schools and homes. He asserted that the tests 'are of no small significance to promote the education and life quality of those who are less equipped by nature. From the standpoint of the teaching personnel, the sorting of children into special schools represents a significant relief in their work'.<sup>33</sup> Vogt's statement is characteristic of a way of arguing between care and control, nature and nurture, which was widespread and took different shapes among Norwegian scholars, physicians, pedagogues, and politicians supporting segregated special education.<sup>34</sup> The broader sentiment in this group leaned towards sympathy for the mental hygiene movement in all its various manifestations, including its intersection with eugenics. This assumed a connection between "mental deficiency" and "social inefficiency", leading to a host of social problems. In line with this thinking, Vogt suggested that intelligence tests could be useful beyond schools, such as in forensic psychiatry and testing young criminals, and possibly in racial hygiene and in controlling marriages between unfit individuals.

---

33 Ibid., 2. This article/interview with Vogt was reprinted in *Skolebladet* 24.50 (1921), 761–62. Earlier the same year, Vogt had published another newspaper article, in which he introduced in detail the practice of intelligence testing and emphasized that well-educated psychologists should perform this task, while the collaboration with pedagogues was necessary, see Ragnar Vogt, 'Maaling av Intelligens', *Morgenbladet*, 64, 26 February 1921. This was again reprinted in the school magazine, *Norsk Skoletidende*, 22 and 26 (1921), pp. 323–25 and pp. 374–75. These two publications at the beginning of the Association's work attest to the desire to educate school people on intelligence testing and gain their support. Vogt's use of the term "psychologist" should be understood as rather loosely referring to those with education in psychology, not necessarily restricted to psychologists. When the psychologist Åse Gruda Skard edited a volume on the first Norwegian psychologists, a chapter was devoted to Vogt and his understanding of psychology 'as a basic subject for a scientifically founded psychiatry', see Per Ancheren, 'Ragnar Vogt 1870–1943', in *Psykologi og Psykologar i Norge*, ed. by Åse Gruda Skard (Oslo: Universitetsforlaget, 1959), pp. 79–80 (p. 79).

34 Here, I agree with Lina Bryne Larsen's observation, who argues that Vogt's language was sharper than that of most teachers, and I would suggest that this was the case with most people coming from the medical side, although their tone depended on their audience; see "'Med Maaling Følger Viden": Intelligensmåling i den Norske Folkeskolen 1900–1945' (Unpublished master's thesis, University of Bergen, 2006), p. 38.

## The Long Process of Standardization

The Oslo Special Schools Association's initial attempt to garner financial support for developing a Norwegian standard was unsuccessful. Those taking the initiative, however, were undeterred, as intelligence testing, in various forms, had already been used in Norway from the end of the nineteenth century. One of the first Norwegian pediatricians, Carl Looft, pioneered the examination of children's *psykiske utviklingsmangler* [mental development deficiencies], mostly based on the examination of children attending *Sæthres Institut for aands svage Børn* [Sæthre Institute for Feeble-minded Children], which also accepted *sinker* [slow learners], and ran special classes in Bergen.<sup>35</sup> Employing tasks he devised himself, the assessment included evaluating their performance in reading, arithmetic, and the Bible, measurements of their heads and other bodily dimensions, along with an exploration of their social environment.<sup>36</sup> The main question was the connection between idiocy or imbecility and other conditions such as epilepsy, malfunctions of the thyroid gland, rickets, infectious diseases, or parents' alcoholism and mental disorders. Looft thought that childhood mental deficiencies were mainly the result of inherited mental degeneration.

Following the full development and international implementation of the Binet–Simon scale, Looft translated it into Norwegian, assessed sixty-seven boys from two ordinary private and public schools, and printed the Norwegian version in a research paper.<sup>37</sup> Intelligence testing and an examination of a child's psychological constitution, Looft asserted, could provide answers to socially and pedagogically important questions and lead to international school reforms. According to Looft, every year, school doctors had to face the same question 'about sorting out "the dead weight"'.<sup>38</sup> Intelligence testing could contribute to a scientifically grounded, medicine-driven classification of abnormal school students into

35 Carl Looft, *Kliniske og Ætiologiske Studier over Psykiske Udviklingsmangler hos Børn* (Bergen: John Grieg, 1897).

36 As John Carson argues, craniometry as part of physical anthropological (and often racial) examinations contributed to establishing the notion that intelligence was both unitary and quantifiable. By the time Looft started his investigations, however, most international researchers were moving away from measuring intelligence in the body and aimed at the mind, see *The Measure of Merit*, p. 109.

37 Carl Looft, 'Intelligensundersøkelser av skolebarn', *Medicinsk Revue*, 10 (1913), 569–85.

38 *Ibid.*, p. 582.

finer categories such as *imbecil* [imbecile] and *debil* [debile] and lead them to the respective training schools for the feeble-minded and special classes.<sup>39</sup> He perceived the method as easy to apply, low in resource use, and rather engaging, providing a holistic picture of a child's intelligence, including 'practical moral intelligence', which was crucial for an individual's functioning in society.<sup>40</sup> Although critical voices were being raised and new methods presented internationally, Looft concluded, referring to the Binet–Simon scale, 'for now it is the best we have'.<sup>41</sup>

Looft's work reached the school administrators in Oslo when they started reforming the school doctor system a couple of years later. Already in early 1915, the committee supervising the school doctor system had received copies of Looft's paper, along with a draft for changes in the system, which did not include intelligence testing, but emphasized the role of school doctors in examining the mental condition of students and advising how they might best benefit from schooling.<sup>42</sup> Soon after, the committee contacted the school inspector in Bergen to find out, among other things, if they used the Binet–Simon scale when deciding on the transfer of children to special schools and how satisfied they were with the system.<sup>43</sup> The correspondence archive shows that the response was quick and encouraging. While Looft had undertaken these examinations with entire school classes for his own interest or in doubtful cases of transfers to special schools, a couple of head teachers had also carried out testing upon Looft's advice.<sup>44</sup> The inspector emphasized that they would strive to apply testing in all schools as part of the school doctor examination, and concluded that 'the Binet–Simon method is an especially trustworthy aid in assessing intelligence'.<sup>45</sup>

---

39 Ibid., p. 582

40 Ibid., p. 584

41 Ibid.

42 Byarkivet i Oslo [hereafter, 'BAO'], Skoleetaten—skolelægeinstitutionens ordning (14 February 1915).

43 Ibid. (17 February 1915).

44 Ibid. (22 February 1915). In a report to his medical colleagues, however, Looft, who was employed by the Bergen school board as a doctor for training special schools and schools for feeble-minded children since 1916, argued that having psychiatric education was an advantage for the overall examination of children; 'Sinkers og Aandsvake Barns Skoleforhold i Bergen', *Tidsskrift for Den Norske Lægeforening*, 45.4 (1925), 198–201.

45 BAO, Skoleetaten—skolelægeinstitutionens ordning (22 February 1915).

Meanwhile, the question of moral intelligence was central in the work on youth delinquency of the Oslo-based psychiatrist Sigurd Dahlstrøm, who sought to investigate the connection between the young delinquent and the neglected child.<sup>46</sup> Beginning in 1912, Dahlstrøm incorporated intelligence testing in his psychiatric examinations of prisoners and patients. For the research leading to his doctoral dissertation, he used the 1911 revision of the Binet–Simon scale, assessing 200 persons in provisional detention. He dedicated around eight years to studying their backgrounds, including how they grew up, their school environment, and their psychopathology. According to Dahlstrøm, the causes behind criminal behavior were both environmental and hereditary, with the first exerting greater influence.<sup>47</sup> He commented that the prisoners found the Binet–Simon examination interesting and amusing, and tried to answer to the best of their ability. Dahlstrøm, however, was less enthusiastic about the scale and its results. He mentioned the changes in specific test items suggested by other international researchers and wondered whether the poor performance of his subjects related to the fact that these tasks were meant for children, or to his sample being “special”. The published thesis included an appendix with the Binet–Simon scale translated into Norwegian, along with brief instructions for each individual test item, and reprints of the pictures and other non-linguistic materials.

Dahlstrøm’s work was already outdated by the time of publication in 1922 due to new attempts to create a Norwegian standard based on the revised and widely circulating 1916 Stanford–Binet scale. Although he had planned to continue his work with intelligence testing with a sample of people with a “normal background”, he never returned to it.<sup>48</sup> Instead, he moved to the position of chief physician at the newly established psychiatric department of Ullevål Hospital. His psychiatric work focused almost exclusively on youth criminality and mental disease, with special interest in mental hygiene, greatly influenced

---

46 Sigurd Dahlstrøm, *Studier over Unge Fanger. 1. Forsøk med Binet–Simons Intelligensprøver* (Kristiania: Aschehoug, 1922).

47 In the years to come, Dahlstrøm wrote and held lectures on the diversity of factors influencing criminal behavior, and often suggested removing children from negative surroundings as early as possible. See for example, Sigurd Dahlstrøm, ‘Arv og Miljø’, *Tidsskrift for Den Norske Lægeforening*, 49.12 (1929), 581–91.

48 Dahlstrøm, *Studier over Unge Fanger*, p. 152.

by his yearly visits abroad, particularly to the United States.<sup>49</sup> In 1930, Dahlstrøm established the Mental Hygiene Association in Norway, but passed away abruptly and dramatically only three years later as he was giving a lecture on mental hygiene.<sup>50</sup>

During the early 1920s, Monrad Jacobsen had also traveled around Europe to study special education institutions, with particular focus on the process of sorting out children.<sup>51</sup> In Germany, she reported that schools were still relying on pedagogical insight for such decisions, using some form of intelligence testing as a supplementary tool, although teachers were interested in developing a standardized form, especially adapted for their country.<sup>52</sup> When she met Théodore Simon, he was adamant that local adaptations of the Binet–Simon intelligence scale were justifiable and the scale ‘was entirely sufficient for determining the child’s placement in school’ without undermining the importance of pedagogical input.<sup>53</sup> After modestly trying the tests within her own class, she concluded that ‘from being skeptical, I have at least become a believer’.<sup>54</sup> Her visit to England was accompanied by a short article in English for the journal of the Central Association for Mental Welfare,<sup>55</sup>

49 Among his various trips, Dahlstrøm received a scholarship by The Rockefeller Foundation ‘to study the American treatment methods of psychopathic children’ in 1926, *Nordmanns-forbundet*, 19.1 (1926), p. 206.

50 Anton Sunde, ‘Sigurd Dahlstrøm in memoriam’, *Tidsskrift for Den Norske Lægeforening*, 53.9 (1933), 453–55.

51 Traveling abroad was an important part of teacher education, and female teachers who were the majority in special schools often received funding for such study trips, as shown in several reports in the periodical *Vor Skole*, a publication of the *Norsk Lærerrineforbund* [Norwegian Women Teacher’s Association]. A special school teacher from Arendal, Marta Sørensen, is a case in point: she was resident for six to seven weeks at Columbia University in 1920, where she observed intelligence testing; see ‘Intelligensprøver med Barn’, *Vor Skole*, 11.37 (1922), 338–40. Most of these female teachers, including Monrad Jacobsen, were unmarried and came from families with higher social status, such as government officials. On the social background of teachers, and especially on the differences between women and men, see Hagemann, *Skolefolk*, pp. 63–87.

52 Julie Monrad Jacobsen, ‘Fra et Besøk ved Tyske Skoler’, *Vor Skole*, 10.20 (1921), 131–33.

53 Julie Monrad Jacobsen, *Særskolen, Dens Maal og Midler* (Göteborg, 1922), p. 17.

54 Ibid., pp. 17–18.

55 The Central Association for Mental Welfare was established in 1913 in connection with the Mental Deficiency Act of 1913 as a service to local authorities, which were overwhelmed by the task of identifying who was subject to the Act in their area. The Act was passed with a large majority and marked mentally defective people ‘as lying outside the boundaries of responsible citizenship: in need of care and control, with the right to be given humane care, but too irresponsible to exert their own civil rights



in which she confirmed that Norwegian standards for the Binet–Simon tests were being developed.<sup>56</sup> In language reminiscent of mental hygiene and eugenics discourse, Monrad Jacobsen wrote, ‘the more the normal children can learn and develop, the greater the difference, the more ours are defeated in the struggle for existence’.<sup>57</sup> Echoing again the mixture of philanthropic care and biological understanding of intellectual disability that Vogt expressed above, she concluded that ‘these stepchildren of society less endowed by nature than others, demand indeed more attention, demand a plus that will cancel their minus’.<sup>58</sup>

### Psychologist Martin L. Reymert’s Unfinished Business

Despite the early, medicine-driven attempts to develop a Norwegian version of the Binet–Simon scale, the Oslo Special Schools Association proposed *Det psykologiske instituttet* [the Department of Psychology] and its director, Professor of Philosophy Anathon Aall, as the most qualified to lead this effort.<sup>59</sup> While psychology was still working to disentangle itself from the influence of philosophy, Aall had, since 1909, managed to establish a center for empirical, experimental, and instrument-driven psychology with several disciples.<sup>60</sup> Having studied both philosophy and experimental psychology abroad, he had previously taken interest in the topic and had already presented an overview of international advancements in intelligence testing to a teacher audience in 1915.<sup>61</sup> Although this was not Aall’s primary focus, he considered the issue to be one of the most important at that time in Norway, and promised his full support. The person who undertook the task was the vice-president of the Association, psychologist Martin Luther Reymert.

Reymert had a background in teacher education, served as a teacher

---

or to have the right to liberty’, see Mathew Thompson, *The Problem of Mental Deficiency: Eugenics, Democracy, and Social Policy in Britain, c.1870–1959* (Oxford: Clarendon, 1998), 37–60 (p. 54). The Act, which was a child of the eugenics movement, introduced the categories of “idiot”, “imbecile”, “feeble-minded”, and “moral imbecile”.

56 Julie Monrad Jacobsen ‘The Special School in Norway and Work Connected with It’, *Studies in Mental Inefficiency*, 15.3(2) (1922), 25–31.

57 *Ibid.*, 30.

58 *Ibid.*, 31.

59 ‘Til Styret for Forskningsfondet av 1919’, 664.

60 *Klinikk og Laboratorium*, pp. 17–43.

61 Anathon Aall, ‘Om en Ny Metode for at Maale Skole Dygtigheten hos’, *Skolebladet* 18.2 (1915), 17–18.

at one of Oslo's special schools from 1914, and earned his doctorate in experimental psychology from Clark University in 1917 under the supervision of the pioneering American psychologist Granville Stanley Hall.<sup>62</sup> After a two-year research stay at the University of Iowa with the influential Swedish-American psychologist and eugenicist Carl Emil Seashore, known for his research on musicology and human development, Reymert returned to Norway. During this period, Reymert worked as a scientific assistant at the Department of Psychology with Aall, taught educational psychology at the Norwegian College of Agriculture, and held courses in child psychology and intelligence testing at the University of Oslo. He wrote in scientific and popular journals on issues related to psychology and education, reviewed books, and established the international journal *Scandinavian Scientific Review*, featuring contributions from philosophy, psychology, and the science of education. Yet, the young scientist was feeling anxious, and was 'longing back to young America', away from Europe, 'where the grandest asset a man can have is to *appear as*, and also to feel himself as damned *old*, as only possible!'<sup>63</sup> In the meantime, he had 'a heap of irons in the fire all the time', with intelligence testing being one of them.<sup>64</sup>

In 1923, Reymert secured funding from several sources to conduct research on intelligence testing. This included 1,500 Norwegian crowns specifically earmarked for studying the systematic examination of schoolchildren in England.<sup>65</sup> The funding was most likely linked to his participation at the VII International Congress of Psychology in Oxford

---

62 Being a student of Hall brought Reymert into a group that included the most prominent psychologists working with intelligence testing, such as Terman and Goddard, who were also students of Hall. The biographical information on Reymert given here is based on a series of biographical sketches and notes found at the Archives of the History of American Psychology—The University of Akron [hereafter, 'AHAP'], Martin L. Reymert Papers, M 2896, Folder 4, Martin L. Reymert.

63 AHAP, Martin L. Reymert Papers, M 1227, Folder G. Stanley Hall, 'Letter to Hall from Reymert', 7 December 1922.

64 Ibid., 7 December 1922.

65 This funding was given by Generalkonsul O. J. Storms stipendiefond, owned and managed by the University of Oslo, with the aim to support younger academics with research scholarships that would bring back knowledge on French and English culture and intellectual life; *Det Kongelige Frederiks Universitets Årsberetning, 1ste Juli 1922–30te Juni 1923* (Oslo: A. W. Brøgger, 1923), p. 239. During the same year, Reymert received additional funding (1000 Norwegian kroner) from the university budget as 'help for the execution of a purely scientific-theoretical part of a larger project on intelligence research', Ibid., p. 241.

in the summer of 1923. The congress opened with a symposium that debated the nature of general intelligence, and at its conclusion, an international committee was formed to promote intelligence testing in all countries, with Reymert as the Norwegian representative.<sup>66</sup> Upon his return, Aall, Vogt, and Dahlstrøm, together with Professor of Philosophy Harald Schjelderup and Professor of Physiology Sophus Torup, composed a to-whom-it-may-concern letter detailing these developments.<sup>67</sup> They expressed their desire for intelligence research to be actively pursued in Norway, highlighting Reymert's systematic international training and preparatory work. They fully endorsed his plan to develop Norwegian standards and hoped for positive support and potential financial backing from the authorities.

Shortly thereafter, Reymert secured additional funding for his research.<sup>68</sup> In a subsequent interview, he emphasized that Norway was the only country lacking its own intelligence tests.<sup>69</sup> He noted the need for both group tests, offering a quick way to identify individuals with lower general intelligence, and the standardization of individual tests. The latter, a more gradual process requiring thorough scientific grounding, was considered a priority. When discussing the significance of intelligence testing for Norwegian conditions, Reymert pointed to the two percent of those deemed mentally defective, particularly "*Morontype*" [the moron type], among whom sexual offenders were frequently found. He specified that the morons were the boundary cases to the normal, and were internationally considered as the most dangerous as they could go unnoticed by society and the authorities. While he agreed with Vogt that harsher punishment would not be beneficial, he advocated for practical solutions utilizing tests to segregate such individuals as

66 *VIIth International Congress of Psychology Held at Oxford from July 26 to August 2, 1923, under the Presidency of Charles S. Myers, Proceedings and Papers*, ed. by Charles S. Myers (Cambridge: University Press, 1924).

67 AHAP, Martin L. Reymert Papers, M 2896, Folder 5, Business Correspondence, and Papers, 1917–1929, 'Letter by Aall, Vogt, Schjelderup, Dahlstrøm, and Torup', 1 September 1923.

68 Reymert received funding from Statens forskingsfond (The Scientific Research Fund of 1919, under a new title) and the private charitable foundation Nansenfondet for 'a large work with intelligence measurement', as well as from Universitets jubileumsfond for 'the preparation of an intelligence study', *Det Kongelige Frederiks Universitets Årsberetning, 1ste Juli 1923 –30te Juni 1924* (Oslo: A. W. Brøggers, 1926), p. 92 and p. 282.

69 'Psykologi og Intelligensprøver i Norge', *Aftenposten*, 619, 25 October 1923, p. 1.

early as possible.<sup>70</sup> In the United States, he noted, several states offered the choice between castration or lifelong incarceration in institutions. These statements reflect not only his Norwegian affiliations but also his American connections with Hall and Seashore, and align with strong eugenicist ideas.<sup>71</sup> The timing of Reymert's interview coincided with debate about a new Norwegian sterilization law, which would allow voluntary sterilization for reasons other than medical ones.<sup>72</sup>

Despite Reymert's reassurances that standardization was underway, there was little evidence of it. In November 1923, he received an offer for a position at Wittenberg College in the United States, complete with his own modest laboratory.<sup>73</sup> However, he hesitated to accept as he did appear to enjoy the trust of his Norwegian colleagues and continued to receive funding from the university for his research.<sup>74</sup> In 1924, the journal of the Norwegian Medical Association briefly mentioned Reymert's presentation titled 'Intelligence tests conducted on Norwegian schoolchildren', held during the meeting of the *Norsk*

---

70 Vogt was one of the foremost experts in forensic psychiatry and held a seat in *Den rettsmedisinske kommisjon* [Norwegian Board of Forensic Medicine]. In 1922, he was a member of the law reform commission, which set the grounds for the establishment of the sterilization law of 1934.

71 While Hall is often remembered for his contributions to the child study movement from the field of psychology, newer literature reveals his racist and sexist views, which placed "savages" at a child-like developmental stage and girls second to boys. For an overview of his theories on the child and critical literature, see Bergstrom-Katz, 'On Intelligence Tests', pp. 44–50. Seashore's equally prominent standing in psychology, his contributions to educational psychology and support for radical educational reforms, has played a role in silencing his hardline eugenicist positions, which recent literature finally discusses; see Julia Eklund Koza, *"Destined to Fail": Carl Seashore's World of Eugenics, Psychology, Education, and Music* (Ann Arbor, MI: University of Michigan Press, 2021). Eklund Koza discusses specifically his role in developing IQ tests for all kinds of institutions and purposes, including identifying musical talent, and his perception of ability as hereditary, hierarchically connected to the white race and males.

72 On the connection between legal reforms, eugenics, and sterilization, see also Haave and Kyllingstad, chapter 4, and Gröning and Skålevåg, chapter 3, in this volume.

73 AHAP, Martin L. Reymert Papers, M 1227, Folder G. Stanley Hall, 'Letter from Reymert to Hall', 26 November 1923. As he later explained to Hall, it was not only being busy in Norway that made him hesitant. The low salary, the religious connections of Wittenberg College, and being in a small place that would be hard to escape, all contributed to his apprehension; AHAP, Martin L. Reymert Papers, M 1227, Folder G. Stanley Hall, 'Letter from Reymert to Hall', 5 March 1924.

74 Reymert continued to receive funding from Statens forskingsfond and Nansenfondet for his work on intelligence testing; *Det Kongelige Frederiks Universitets Årsberetning, 1ste Juli 1924 –30te Juni 1925* (Oslo: A. W. Brøgers, 1926), p. 84.

*Psykiatrisk Forening* [Norwegian Psychiatric Association] but never published.<sup>75</sup> His revision of the Stanford–Binet appeared in the same journal issue, directed especially for use by asylums and forensic psychiatry experts.<sup>76</sup> This was an advertisement for the whole scale set, priced at 48.5 Norwegian crowns, approximately equivalent to 1,350 Norwegian crowns today. It comprised thirty typed pages of instructions reviewed by Vogt and Schjelderup, a set of pictures and scoring cards, twenty-five record booklets, a wooden formboard, and a set of five standardized weights.<sup>77</sup> Examination of the record booklet reveals no indication of collaboration with Houghton Mifflin, the company that copyrighted the Stanford revision in 1916.<sup>78</sup> The only visible stamp is from Arthur Nelson’s printing house. Within Reymert’s archive, there is no information regarding negotiations or discussions with Lewis Terman, despite Reymert having an extensive American network and later being in contact with Terman regarding research findings on the effects of the environment on intelligence.<sup>79</sup>

Some of the most intriguing differences between the Stanford revision and Reymert’s version appear on the first page of the test booklet, where the tester should make general notes.<sup>80</sup> The American version included a series of physical examinations, such as standing and sitting height, weight, head circumference, right and left grip, lung capacity, markers of disease, physical defects, and disease history.<sup>81</sup> However, the Norwegian version omitted all these elements, except for walking and talking age. It remains unclear whether this omission was accidental or a deliberate departure from earlier physical anthropological and medical approaches in favor of psychological examination, as the medical record was noted

75 ‘Norsk Psykiatrisk Forening’, *Tidsskrift for Den Norske Lægeforening*, 17–18 (1924), 840–41.

76 ‘Intelligensprøver til Bruk Særlig for Asyler og Rettspsykiatriske Sakkyndige’, *Tidsskrift for Den Norske Lægeforening*, 17–18 (1924), 865–66.

77 Unfortunately, Reymert’s translated test manual has not been located, which is not surprising as it was probably printed in a limited number and privately sold.

78 Martin L. Reymert, *Stanford Revisjon av Binet–Simons Intelligensprøver: Forsøk paa Tilpasning til Bruk for Norsk Standardisering* (Oslo: Arthur Nelson, 1924).

79 Reymert and Terman corresponded between 1938 and 1939 regarding a five-year study by the former on the possibility to improve IQ in a superior environment, AHAP, Martin L. Reymert Papers, M 1229, Folder Miscellaneous Correspondence.

80 Reymert, *Stanford Revisjon*, p. 1.

81 Lewis M. Terman, *Record Booklet for the Stanford Revision of the Binet–Simon Tests as Described in Terman’s The Measurement of Intelligence* (Boston, MA: Houghton–Mifflin, 1916), p. 1.

separately by each institution's doctor in any case. The relationship between intelligence and class, which was a core issue in international debates from early on, particularly in relation to possible environmental effects, also appeared differently.<sup>82</sup> In the Norwegian case, this was expressed through noting the social position of the father, most probably through occupation, while the American version referred to categories of social status, ranging from very inferior to very superior.<sup>83</sup> Another important departure from the US booklet, which noted if a child was considered superior to their classmates and had the option to skip a class, was omitted from the Norwegian version as this option was rather uncommon at the time, if it existed at all. In both versions, the teacher's qualitative classification of the child's intelligence was recorded. The rest of the booklet, with minor alterations such as in the case of coin names, was a verbatim translation—although with several traces of sloppiness—from the Stanford revision. Whether this version was put into work for standardization, and what its concrete results were remains unclear, although Reymert wrote in his professional biography that he had standardized intelligence tests for Norway between 1921 and 1925. In his attempt, standardization seems to imply stagnation rather than change for the test and the person. He found himself stuck with this task in his home country for five years, lacking better choices.

Hall died in April 1924, before being able to assist Reymert with the negotiations with Wittenberg College. However, it appears that Reymert was successful in achieving his goals, and after the summer of 1925, he permanently returned to the United States. Institutions and individual doctors were encouraged to purchase the revised test materials before his departure, as obtaining them later at cost price would be difficult.<sup>84</sup> Psychologist Åse Gruda Skard noted, in an

82 The relationship between intelligence, class, heredity and environment dates back to the debates of the second half of the nineteenth century with works such as Francis Galton's and his Victorian contemporaries, see for example Ageliki Lefkaditou, "'This Wonderful People': Darwin, the Victorians, and the Greeks", *Journal of Modern Greek Studies*, 36.1 (2018): 97–124.

83 The relationship between fathers' occupation and children's intelligence was a research interest Reymert worked on later and argued that the children of skilled laborers seemed to have superior intelligence when compared to unskilled workers, see Martin L. Reymert and John Frings, 'Children's Intelligence in Relation to Occupation of Father', *American Journal of Sociology*, 41.3 (1935), 351–54.

84 'Direktiver og Instrumentarium til Intelligensmaaling', *Tidsskrift for Den Norske Lægeforening*, 11 (1925), 598.

otherwise positive portrayal of Reymert, that 'he started many things which remained unfinished'.<sup>85</sup> An example she cited was the 'Army Alpha' group test that he never managed to standardize, although he sold it alongside the Stanford–Binet test materials. As a final act in this long road to standardization, he successfully proposed that 30,000 Norwegian crowns from the financial gains of the *Det norske pengelotteri* [Norwegian lottery] be allocated to the standardization of intelligence tests and related purposes.<sup>86</sup>

### An Accidental Tester: Psychiatrist Johan Lofthus Takes the Helm

In April 1923, Johan Lofthus became the first doctor specializing in psychiatry to work within special schools following a petition to the municipality of Oslo by Ragnar Vogt, head school doctor and health inspector for Oslo, Carl Schiøtz, and school inspector Olav Schulstad.<sup>87</sup> After earning his medical degree in 1913, Lofthus had acquired experience working with psychiatric issues within various temporary and part-time positions. He served as reserve doctor at the Sanderud psychiatric institution (1918–19) and the psychiatric department at Ullevål Hospital (1920–23).<sup>88</sup> From 1921, he worked as a police and prison doctor, forensic expert, and physician in Kristiania county to decide on the care and treatment of mentally ill individuals. The position at special schools was initially limited to two hours per day, later prolonged to four,<sup>89</sup> which allowed Lofthus to keep his

---

85 Åse Gruda Skard, 'Martin L. Reymert 1883–1953', in *Psykologi og Psykologar i Norge*, pp. 100–102 (p. 102).

86 The profits from the state Norwegian lottery were to be invested in a host of socially beneficial purposes, from medical ones to cultural activities. The rationale behind the decision to support intelligence testing is not known.

87 The psychiatrist Wilhelm M. Hornemann, at the time reserve doctor at the psychiatric department of Ullevål Hospital, was offered the position, which he rejected. When the offer was made to Hornemann, he was on his way to a higher position as head doctor and director of Blakstad psychiatric asylum. BAO, *Aktstykker Kristiania kommune*, Dokumenter II. 2, 1922–23, Dokument nr. 15, p. 10.

88 Biographical information on Lofthus comes from *Studentene fra 1907: Biografiske Oplysninger Samlet til 25-Års-Jubileet 1932* (Oslo: Grøndhal & Sønns, 1932), pp. 248–49. This information was self-reported, therefore it has been cross-checked with other sources.

89 BAO, *Aktstykker Kristiania kommune*, Dokumenter II. 2, 1926–27, Dokument nr. 15, p. 38.

private practice specializing in nervous and mental disorders. From 1924, he also worked as a school doctor for ordinary schools in Oslo, as Vogt's assistant in pathology and anatomy at Gaustad psychiatric asylum, and as expert assistant to the head physician of the mental health services.<sup>90</sup>

In this intricate network of minor positions, Lofthus's main responsibility was to advise on the transfer of schoolchildren from ordinary to special schools, in addition to his duties as general doctor for the two special schools in Oslo.<sup>91</sup> Intelligence testing played a crucial role in this segregation process. Lofthus, lacking any prior experience with such methods or school-related work, was obligated to travel to neighboring Scandinavian countries and learn from their examples.<sup>92</sup> For a brief period in the autumn of 1923, he visited Stockholm to consult with special school doctor Dr Alfild Tamm, and to Copenhagen to meet school doctor Dr. H. P. T. Ørum and Professor Christian Keller.<sup>93</sup> All three had contributed to introducing intelligence testing to their respective countries and were actively involved in the controversies and practices around the assessment, treatment, and control of those characterized as mentally defective. Upon returning to Norway, Lofthus began using Dahlstrøm's translation of the original Binet–Simon scale, until Reymert's revision became available.<sup>94</sup>

90 See, *Studentene fra 1907*, pp. 248–49.

91 Detailed reports on the responsibilities and initiatives of all doctors working in schools can be found in BAO, Helserådet, 05 Rapporter, Skolelegenes årsberetninger, 1922–40.

92 Johan Lofthus, 'Skolepsykiaterens Arbeid', in *Oslo Særskole*, pp. 74–75 (p. 74).

93 Ibid., p. 74. Alfild Tamm was a complex figure, often remembered as the first female psychiatrist in Sweden, with a great interest in child and youth psychiatry, and the one who established psychoanalysis in the country. She was among those who introduced intelligence testing in the sorting of schoolchildren and collaborated closely with schools to assist teachers in their work of identifying the weaker students; see Thom Axelsson, 'De Svåruppfostrade Barnen: Skolpsykiatrins Framväxt Och Etablering i Sverige 1910–1955', *Scandia*, 86.2 (2020), 60–82. Christian Keller headed the infamous Keller institutions for people with intellectual disabilities in Denmark, with special interest in women thought to be both mentally and morally deficient; he was a known eugenicist and hereditarian; see, Birgit Kirkebæk, 'Sexuality as Disability: The Women on Sprogø and Danish Society', *Scandinavian Journal of Disability Research*, 7.3–4 (2005), 194–205.

94 Lofthus, 'Skolepsykiaterens Arbeid', p. 75.





Fig. 1.1 Portrait of psychiatrist Johan Lofthus (c.1935). He worked with the adaptation of the Stanford-Binet scale in Norway after he obtained the position of school psychiatrist in Oslo in 1923. He was a staunch ambassador for the use of intelligence tests in schools, prisons and mental institutions and collaborated closely with school staff, doctors and psychologists to advance this cause. After the publication of his standard, known as *Osloprøvene* [Oslo Tests], he left his position. Photographer unknown/Oslo Museum, CC0 1.0.

Less than a year later after assuming the position in schools, Lofthus had his first opportunity to present his work on intelligence testing at the congress for Physical Education and School Hygiene held in Oslo in the autumn of 1924.<sup>95</sup> Interest in school hygiene had been widespread in Norway since the early twentieth century, influenced by developments in the rest of Europe and the United States.<sup>96</sup> The general principle of ‘a healthy mind in a healthy body’ brought together a diverse group of people with varying attitudes towards the roles of environment and heredity, all aimed at carefully engineering the school environment in all its aspects. The congress garnered significant public attention with 300 participants, including prominent Norwegian progressive educators and educationalists such as Anna Sethne and Otto Grennes, and physicians and hygienists like Schiøtz and Vogt. The programme encompassed discussions on physical exercise, nutrition, sleep, dental and other bodily hygiene alongside new approaches in teaching, open air schools, and a better balance between school classes and everyday life. In his passionate opening plenary titled ‘Health No. 1, Teaching No. 2’, Carl Schiøtz emphasized the need to transition to an active understanding of school hygiene, including care for the body, to make teaching results ‘much more fruitful’.<sup>97</sup> The question of segregation in

95 Johan Lofthus, ‘Intelligensundersøkelser og Psykiatrisk Arbeide ved Kristiania Komm. Skoler’, in *Nordisk Kongress for Fysisk Opfostring og Skolehygiene i Kristiania 28. septbr.–1. Oktober 1924*, ed. by Otto Grennes (Eberh. B. Oppi forlag: Oslo, 1925), pp.124–31.

96 At the dawn of the twentieth century, which Swedish feminist writer Ellen Key prophesied would be the century of the child, emerged the movement of child science, or pedology, in several western societies. Its aim was to study children’s development and behavior as a multidisciplinary field encompassing pedagogics, hygiene, psychiatry, psychology, and criminology. For a case in point, highlighting the intersections between pedology, hygiene, racial studies and criminology, see Angeliki Lefkaditou, ‘Naturalizing the Nation: Physical Anthropology in Greece, 1880s–1950s’ (Unpublished PhD thesis, University of Leeds, 2021), pp. 204–28. In Norway, this movement became popularized under the guise of school hygiene, and with the strong leadership of physicians, who enjoyed high social prestige and political influence in matters of social policy.

97 Carl Schiøtz, ‘Helsenr. 1, undervisningenr. 2’, in *Nordisk Kongress for Fysisk Opfostring*, pp. 8–24 (p. 14). Schiøtz was at the time already an authority on matters of community health and had conducted extensive examinations of schoolchildren’s physical development in relation to nutrition. His role in the school hygiene movement was pivotal, and in the 1930s, he was behind the initiative to introduce the healthier “Oslo breakfast” to all schools in Oslo; see Astri Andresen and Kari Tove Elvbakken, ‘From Poor Law Society to the Welfare State: School Meals in Norway 1890s–1950s’, *Journal of Epidemiology and Community Health* (1979), 61.5 (2007), 374–77.

this new, progressive and useful school was central and manifested in a teaching demonstration of a special class for weak students. Lofthus's talk on intelligence testing, which received a special mention by Schiøtz as a contribution to 'a chapter marked by rewritings and controversy', further reflected this overarching novel direction.<sup>98</sup>

Lofthus began his presentation by addressing the resistance encountered when deciding to transfer children to special schools, particularly from parents who often accused teachers of attempting to remove troublesome pupils. Nevertheless, he explained why such transfers were necessary for the benefit of the entire class and individual children who did not progress with ordinary teaching. In such cases, pedagogical judgment did not seem entirely satisfactory, but 'a measuring method for mental abilities, which could be expressed in numbers, well, that was something quite more sober and indisputable'.<sup>99</sup> He cautioned that such mathematical precision in the measurement of intelligence was not entirely attainable at the time, but the tests in combination with the teacher's judgment could provide a solid foundation for action. This consideration for teachers' perspectives remained a consistent theme throughout Lofthus's work and in the interactions within the group of doctors, psychologists, and teachers associated with special schools. In any case, the entire transfer process required collaboration between school authorities and the school psychiatrist. A transfer was not allowed without the psychiatrist's recommendation, while in cases of disagreement between the head teachers of the ordinary and the special school, the psychiatrist's suggestion was decisive.<sup>100</sup>

While Lofthus remarked that he wished he could conduct a thorough examination of everyone, particularly of their home environment, time constraints compelled him to rely on psychiatric examinations and intelligence testing.<sup>101</sup> Consequently, the intelligence quotient (IQ) score became the primary criterion for determining who fell within the range of normality during the psychiatric assessment. Lofthus found Terman's

98 Schiøtz, 'Helsen nr. 1, undervisningen nr. 2', p. 21.

99 Lofthus, 'Intelligensundersøkelser', 124–25.

100 Ibid., p. 125. If the disagreement persisted, the head school doctor and the school inspector would handle the case; BAO, Skoleetaten, Skolestyrets arkiv, 1913–31, Mappenr. 132, 'Regler for Overflytting sv Elever til Særskolen'.

101 Lofthus, 'Intelligensundersøkelser', pp. 124–25.

suggestion of an IQ between 90 and 110 as normal too restrictive. He preferred starting lower, allowing children with an IQ above 80 to remain in ordinary schools. His focus was on identifying ‘the children who rightfully belonged to the special school’, specifically those with an IQ under 70.<sup>102</sup> Questions such as, ‘How low should one go? Should special schools be inundated with low-standing idiots?’ followed, but Lofthus left this lower limit open, likely due to the need for intervention in the highly debated issue of defining and treating those with intellectual disability.<sup>103</sup> Nevertheless, he observed that special schools were admitting children with lower capacities who belonged to training schools for the feeble-minded. Conversely, the lack of special schools in the countryside led to overcrowding the state-owned institutions for the feeble-minded in the towns with children who had higher abilities, leaving the ones who really needed a place there without any option other than staying at home.

Turning to the example of Stockholm, where children with IQs between 77 and 83 were placed in classes that aimed to help them, while all others with an IQ below 77 went to boarding schools for the feeble-minded, Lofthus suggested a new solution for Oslo. He recommended establishing ‘help classes for those who do not lag behind so much. IQ 75–85?’, while ‘[t]hose who score under 75 (70?) can the special school take? The normal class must be freed from *sinkerne* [the slow learners]’, he wrote.<sup>104</sup> Within just a couple of pages, the finer differentiation of schoolchildren at the borderland between those seen as “normal”, “simply slower”, and “mentally defective” changed. As the cutoff IQ values moved up and down between the United States, Stockholm, and Oslo, and even within Lofthus’s own suggestions, the consequences for those tested were also in flux.

This congress presentation marked a programmatic starting point for Lofthus’s further work. In the aftermath of the congress, a core group of participants began touring the country to advance the agenda of school hygiene. Media coverage, both in popular press and professional journals, was overwhelmingly positive, with headlines hailing it as a

---

102 Ibid. p. 126.

103 Ibid. p. 127.

104 Ibid. p. 128.

historical moment in Norwegian school history.<sup>105</sup> Newspaper reports depicted packed rooms with hundreds of teachers, doctors, and parents expressing optimism for this new era in education.<sup>106</sup> Over the following years, Lofthus traversed the entire country to present his work with intelligence testing and special schools. He conducted courses and demonstrations for teachers and contributed numerous articles and interviews. The angle of his talks gradually shifted from introducing testing as a new scientific tool to advocating it as part of essential school reforms.

Towards the end of the 1920s, intelligence testing became intertwined with the broader mental hygiene movement, reflecting Lofthus's growing interest in preventing youth criminality and engaging in anti-alcohol campaigns. Throughout this period, Lofthus cultivated a robust alliance with educators, particularly those within special schools and school administration, and their national and international organizations. This symbiotic relationship harmonized progressive pedagogical thinking, often accompanied by radical ideas on women's roles in society and sexual education, with renewed trust in scientific methods and numerical data.<sup>107</sup> Similar to other social movements, notably eugenics, the progressive educational movement attracted individuals from diverse political backgrounds.

The yearly reports from Lofthus to the head of school doctors, Schiøtz, provide insight into his work as a school psychiatrist. At Sarsgaten's school, he initially worked in a limited space shared with other employees, situated along a bustling corridor that often disrupted psychiatric examinations due to the noise.<sup>108</sup> Eventually, he secured his own office in the loft. At Ullevålveien's school, he even managed to arrange a waiting room.<sup>109</sup> These spatial arrangements were necessary, as

105 Egill Kobbestad, 'Fylkesmøte for Fysisk Opfostring og Skolehygiene', Østlandet, 282, 08 December 1924, p. 1.

106 See for example, 'Det Store Fylkesmøte i Elverum for Fysisk Opfostring og Skolehygiene', Østlendingen, 83, 07 November 1924, p. 1 and 4.

107 Thuen describes vividly how the coming of progressive, reform education in 1920s Norway resulted in less instruction and homework, more activities, less focus on evaluating performance at least during the early school years, and the inclusion of new teaching materials, such as films and slides, see *Den Norske Skolen*, pp. 116–24.

108 BAO, Helserådet, 05 Rapporter, 052 Skolelegenes årsberetninger for skoleåret 1922–23.

109 BAO, Helserådet, 05 Rapporter, 052 Skolelegenes årsberetninger for skoleåret 1926–27

Lofthus observed that relatives frequently accompanied the children 'to provide information about the child's medical history, family relations, domestic circumstances, and its environment outside of school'.<sup>110</sup> Lofthus noted that as his workload increased, meeting more than 150 children per year, so did the interest of both teachers and parents in psychiatric examinations and intelligence testing, particularly as they became more familiar with the schools.<sup>111</sup> Issues of consent became more apparent in physical and psychiatric examinations of graduating special school students, a practice not regulated by law.<sup>112</sup> Lofthus reported that all students could undergo examinations, including intelligence testing, to receive advice on their future endeavors, but participation was strictly voluntary. These assessments were conducted at the request of the children and their parents.

Criticisms of both school segregation and testing appeared sporadically, particularly from teachers at ordinary schools that opposed school segregation.<sup>113</sup> In their reports from abroad, teachers debated the advantages of alternative educational systems, including the use of intelligence testing to determine class placement, but the scientific validity of the tests does not appear to have been a much-debated issue.<sup>114</sup> When Lofthus commented on the criticism intelligence tests

---

110 BAO, Helserådet, 05 Rapporter, 052 Skolelegenes årsberetninger for skoleåret, 1927–28

111 BAO, Helserådet, 05 Rapporter, 052 Skolelegenes årsberetninger for skoleåret 1927–28; BAO, Helserådet, 05 Rapporter, 052 Skolelegenes årsberetninger for skoleåret 1928–29.

112 BAO, Helserådet, 05 Rapporter, 052 Skolelegenes årsberetninger for skoleåret 1928–29.

113 An interesting example of this skepticism is expressed in an article by teacher Hjørdis Saarheim, who was an enthusiastic supporter of the progressive Dalton plan in education, aiming at tailoring teaching to each student's needs, interests, and abilities. While she did not oppose intelligence testing, Saarheim maintained that differentiation according to mental abilities was unnatural as it created an environment in schools lacking the diversity one encounters in the rest of society; see 'Differentiering i Folkeskolen', *Norsk Skoletidende*, 10 (1930), 127–28. At the end of a lecture to school administrators by Lofthus, several members of his audience pointed out that some teachers and parents were prejudiced towards intelligence testing and special education, but noted that it was time to move towards specially adapted teaching and break away from the rigidity of aged-based classrooms; see Johan Lofthus, 'Om Intelligensmåling og Særskoleundervisning', *Norsk Skoletidende*, 38 (1927), 514–17 (pp. 516–17).

114 A thorough examination of teachers' positions towards intelligence testing remains to be conducted, however, I agree with Larsen's conclusion that during the 1920s the teachers were mainly recipients of dissemination about the method rather than active

had received as ‘completely useless’ from Danish school administrators, he responded that ‘either there must be something wrong with the tests used in Denmark, or it must be those who administer the tests, who do not know their stuff’.<sup>115</sup> He added that mastering test administration was difficult even for those with psychiatric experience, but in his work with children from ordinary schools, he never found a contradiction between his results and the judgment of the pedagogues. This was confirmed to the newspaper by all teachers at Møllergaten school, where Lofthus had tested 210 children, although the headmaster noted that he had not yet seen a written presentation of the research results.<sup>116</sup> However, this was about to change, as in 1931, Lofthus published his version of the Stanford–Binet intelligence scale.

### Making the First Norwegian Full-Scale Adaptation of the Stanford–Binet Scale in Oslo

The Oslo tests, or *Osloprøvene*, as Lofthus’s revision came to be known, was so named because the whole process of adaptation took place in the capital with Lofthus as the sole researcher involved in it.<sup>117</sup> Reflecting on this work years after its completion, he commented that he had started with Dahlstrøm’s and Reymert’s translations but ‘none of these were standardized for Norwegian children’.<sup>118</sup> Various issues emerged—some items were deemed inappropriate, others misplaced within age groups, and some appeared useless due to divergent responses between boys and girls. Lofthus committed to a comprehensive revision of the scale, a substantial effort that formed the basis for his doctoral research and paved the way for a coveted permanent position as mental health inspector for the Oslo municipality. The tangible outcome materialized in a hardbound publication of the entire doctoral thesis by Aschehoug, the University-of-Oslo-approved publisher. While Lofthus’s archive

---

users, with few critical voices; see “‘Med Maaling Følger Viden’”, p. 43.

115 ‘Intelligens og Evner i Skolen’, *Dagbladet*, 45, 22 February 1930, p. 12.

116 *Ibid.*, p. 12.

117 While the title “*Osloprøvene*” was sometimes later used to refer to group tests for schools in the Oslo area, in these earlier years it referred to the version of the Binet–Simon test included in Lofthus’s work; *Intelligensmåling* (Oslo: H. Aschehoug & Co, 1931).

118 Lofthus, ‘Skolepsykiaterens Arbeid’, p. 75.

remains elusive, a handful of surviving cards from the test kit, produced based on his revision, offer a glimpse into his work.<sup>119</sup> As with Reymert, there is no evidence of copyrighted material or communication with Lewis Terman in the latter's extensive archive.

The initial and enduringly controversial step in standardizing the Stanford–Binet version involved using a random sample of schoolchildren. Lofthus acknowledged the impracticality of undertaking this task as a lone researcher and pragmatically opted to define a representative sample comprising 150 pupils from Møllergaten elementary school and 50 pupils from Borgerkolen lower secondary school, where most of the first would continue their education.<sup>120</sup> As he reported to the head school doctor Schiøtz, he needed to ‘take as many as possible detailed examinations of normally gifted children in different ages to acquire the necessary ability to judge the mentally defective as correctly as possible’.<sup>121</sup>

The choice of school holds significance for two key reasons. Firstly, Lofthus sought a school and children perceived as unexceptional in both abilities and home situations, a determination he validated with assistance from the Oslo school inspector. While other urban schools—situated in both the economically challenged east and the affluent west areas of the capital—experienced significant population changes, including an influx of children from private schools to public ones, Møllergaten school remained unchanged in the heart of Oslo, with assumed stable socioeconomic variables. Secondly, Lofthus aimed to support this somewhat subjective choice by referencing the work of Carl Schiøtz on the physical development of children and youth, based on the examination of around 10,000 individuals.<sup>122</sup> He documented the average weights of children in each school from Schiøtz's report and examined how this correlated with transfers to special schools, ultimately concluding that Møllergaten appeared genuinely average. As he clarified, ‘the possibility that average intelligence and average social conditions coincide is, however, significant. And this possibility

---

119 The cards are deposited at the Library of the University of Oslo in an undated folder titled *Binet–Simon: Osloprøven, Psykologisk institutt, Oslo*.

120 Lofthus, *Intelligensmåling*, p. 5–12.

121 BAO, Helserådet, 05 Rapporter, 052 Skolelegenes årsberetninger for skoleåret, 1927–28.

122 Lofthus, *Intelligensmåling*, p. 5–12.



increases when a physical property such as children's weight in the area is also almost average'.<sup>123</sup> Lofthus's hypothesis aligned with the prevailing contemporary notion that mental abilities adhered to the same principles as other physical properties, and when appropriately measured, followed a normal distribution curve.

As part of his investigations, Lofthus examined whether the results of intelligence testing of these 210 children would yield the desired bell curve.<sup>124</sup> With a certain satisfaction, he wrote that his curve closely resembled Terman's, although the left side, with lower IQ scores, declined more abruptly, likely due to some of the weakly gifted children already attending special schools. With this discrepancy in mind, we could explain why Lofthus chose to decidedly eschew his rationale for a representative sample by including 25 children under school age from Lunden orphanage, 115 adult psychiatric patients and remand prisoners, and 896 children from Oslo's schools, mostly assumed to be weakly gifted and referred to him for examination. His goal was twofold—to expand his sample size and to correct for the lack of lower IQ scores among the original ordinary schoolchildren sample. Lofthus's work is filled with such statistical acrobatics, ranging from varying sample sizes for each age group to varying years between repeated testing to assess IQ constancy.<sup>125</sup>

For the latter, he re-tested 221 children from the sample of 1,246 at intervals ranging from one to six years. Contrary to accepted orthodoxy, he concluded that IQ scores were not constant toward the lower ends and deteriorated as the years went by. It is important to note that for Lofthus, this finding did not undermine the overall constancy of IQ or his view on intelligence as an inborn capacity. Like-minded psychiatrists and psychologists shared the belief that a person was born with a certain amount of nerve cells, responsible for cognitive capacities through complex processes. These capacities could not be altered beyond certain biological limits, although they could be nurtured or hindered. 'There is nothing more resistant to and unaffected by the environment than the hereditary material', as he answered a journalist inquiring about the

---

123 *Ibid.*, p. 7.

124 *Ibid.*, p. 31.

125 *Ibid.*, pp. 32–42.

inheritance of feeble-mindedness.<sup>126</sup> Testing for intelligence, regardless of the term's exact definition and complexity, was thus crucial to provide a child the opportunity to thrive based on their inborn capacities.

Lofthus's revisions were notable, both in relation to the Stanford–Binet scale by Terman and Reymert's version.<sup>127</sup> From the latter, he departed significantly, retaining only select linguistic choices. Among the major revisions in the Stanford–Binet, Lofthus introduced a novel approach to the vocabulary test, challenging Terman's assumption that selecting words from an 18,000-word dictionary sufficed.<sup>128</sup> Instead, Lofthus opted for the last word on every tenth page from a widely used 60,000-word dictionary. He excluded antiquated words, and after testing children and analyzing how the results correlated with overall intelligence performance, he reordered the chosen words twice. Another significant change was the inclusion of an alternative size-weight illusion test in three age groups, developed by Belgian physiologist and doctor Jean Demoor.<sup>129</sup> Lofthus first encountered this test in 1920 during a study trip to Hamburg. It was part of a series of tests designed to assess feeble-mindedness used by Professor Wilhelm Weygandt, a psychiatric asylum director and racial hygienist. The reason for this inclusion, other than that Lofthus thought of it as a good test, is not obvious.

For the rest of the scale, the most important changes involved moving test items to the age group where sixty-five to seventy-five percent of the sample responded correctly, and eliminating items that showed significant differences between the sexes, which were attributed to differences in practice or environment.<sup>130</sup> For example, Lofthus removed completely from his revision the test for five years requiring the naming of four colors, as he observed a substantial discrepancy in how boys and girls answered, with the latter achieving better results, which he attributed to their 'greater

---

126 'Aandssvakhet Hyppigst på landsbygden både i England og i Norge', *Arbeiderbladet*, 45, 16 October 1930, p. 3.

127 Lofthus's version is compared with Lewis M. Terman, *The Measurement of Intelligence: An Explanation of and a Complete Guide for the Use of the Stanford Revision and Extension of the Binet-Simon Intelligence Scale* (Boston: Houghton Mifflin Co., 1916).

128 *Ibid.*, pp. 136–50.

129 *Ibid.*, pp. 151–61.

130 While differences in intelligence between sexes was a controversial issue for decades, by this time there was almost agreement that there was no biological difference in average intelligence or its distribution between the two sexes. See Terman's summary of the discussion in *The Measurement of Intelligence*, pp. 68–72.

interest in colors'.<sup>131</sup> Terman had also identified this difference, but chose to retain the item, considering it overall excellent despite being likely influenced by home environment.<sup>132</sup> Another example pertains to the test for seven-year-olds, which involved tying a bow knot. Lofthus removed it, upon observing that girls mastered it at a much younger age, and accepted Terman's explanation that girls might be more familiar with bows as part of their clothing.<sup>133</sup> In addition to sex, Lofthus explored differences in performance based on social status and geography. A notable case was the test titled "naming four coins" for six-year-olds, which Lofthus believed depended heavily on the home environment and the child's use of money in daily life.<sup>134</sup> Yet, he retained this item but adjusted the age group.<sup>135</sup> Lofthus also suspected variations in responses between children from rural and urban environments, exemplified by the test item for three-year-olds requiring the recitation of one's surname. Using one's surname was more common in towns, but as he had not conducted research on this matter, he retained the item in its current position in the scale.<sup>136</sup> This is noteworthy, as it points to the scale being an Oslo scale, or at least adapted for bigger towns, although Lofthus already knew that its use would likely be wider.

The changes in Lofthus's revision affected over fifty percent of the items in the scale, although the overall length of the scale remained the same. He consistently compared his findings and item placement with the original Binet scale, Terman's revision, Dahlström's choices in language and placement, as well as alternatives from figures like Burt and the American psychologist Arnold Gessel.<sup>137</sup> Lofthus often questioned Terman's judgment regarding the categorization of answers

---

131 Lofthus, *Intelligensmåling*, pp. 93–94.

132 Terman first remarked that girls 'are somewhat superior to boys in this test, due probably to a greater natural interest in colors', while he did not explain what the home influence could be, see *The Measurement of Intelligence*, p. 165.

133 Lofthus, *Intelligensmåling*, pp. 101–02.

134 *Ibid.*, 99.

135 While Terman attributed the better performance of children from higher classes 'to a superiority in original endowment', Lofthus's position is not clear. However, one may assume that considering intelligence an inborn quality is consistent with keeping this test despite being sensitive to social differences. See Terman, *The Measurement of Intelligence*, pp. 72–73.

136 Lofthus, *Intelligensmåling*, p. 89.

137 Reymer's contribution is not mentioned in Lofthus's revision, although he later referred to it.

as inferior or superior, the influence of schooling or social environment, the rationale for including certain items, or the cumbersome wording of others. The final version of the Oslo scale differed significantly in the answers suggested as acceptable to the test users, but the administration process remained a verbatim translation of Terman's manual. This included acknowledging that there were no pure tests of intelligence, as intelligence was deemed 'all too complicated to be weighted like a bag of flour by anyone who can read figures'.<sup>138</sup> Both Terman and Lofthus agreed that 'the mechanical perfection of the tests can never replace good judgment and psychological insight'.<sup>139</sup> All changes aside, Lofthus's version retained its identity as a Stanford–Binet scale, and potential shortcomings in individual items seemed to be superseded by the system as a whole, still assumed to effectively capture intelligence.

Following the Norwegian academic tradition of a public doctorate defense, Lofthus presented his thesis in a room full of people interested in hearing about intelligence testing at the end of September 1931.<sup>140</sup> The two examiners were probably the most influential psychiatrists of the period, Ragnar Vogt—once a staunch supporter of Lofthus and intelligence testing—and Hans Evensen, the director of Gaustad psychiatric asylum. They were both unimpressed and skeptical whether the thesis was worthy of the doctorate degree. They commented that it was building on a thin empirical basis, was unoriginal and too mechanical—a mere uncritical translation of Terman's work, offering little reflection on the concept of intelligence. Evensen found it difficult to understand Lofthus's enthusiasm for the scale and recalled consulting Simon on a murder, without being impressed by the man. He also noted the fact that Simon had stopped using the tests himself.<sup>141</sup>

Despite these criticisms—fair or not, simply performative or real—the two psychiatrists acknowledged that Lofthus had put great effort in this work, and he was awarded the degree. His standardization was not perfect, but good enough to do the necessary work, as was the

138 Lofthus, *Intelligensmåling*, p. 172; Terman, *The Measurement of Intelligence*, p. 136.

139 Lofthus, *Intelligensmåling*, p. 172; Terman, *The Measurement of Intelligence*, p. 136.

140 Several newspapers reported in detail from the event. See for example, 'Intelligens i Universitetets Gamle Festsal', *Dagbladet*, 226, 28 September 1931, pp. 2 and 9; 'Hva Er Intelligens?', *Aftenposten*, 488, 28 September 1931, p. 2; 'En Mann som Vil Måle Vår Intelligens', *Arbeiderbladet*, 265, 29 September 1931, p. 11.

141 'Intelligens i Universitetets Gamle Festsal', p. 9.

argument for most tests. As Dahlstrøm wrote in his review of Lofthus's book, he and his colleagues were already using the revision and they would continue to do so until other and better tests appeared.<sup>142</sup> Indeed, Lofthus's revision remained in use until the early 1960s, when his successor, psychiatrist Augusta Rasmussen, left special schools, and both the scale and position were substituted by new institutions. Lofthus had continued his career as mental health inspector for Oslo and member of various psychiatric and hygienic committees, including the Norwegian Sterilization Board, since 1933.<sup>143</sup> Nevertheless, he retained a connection to schools as a school doctor, and worked further on standardizing the Dearborn group test for all first-class schoolchildren together with Rasmussen and school inspector Bernhof Ribsskog.<sup>144</sup>

## Concluding Incidents and Thoughts

In March 1929, a government employee and father of a seven-and-a-half-year-old boy wrote to the Oslo School Board requesting his son's transfer to Ullevålveien's school.<sup>145</sup> Hans Eng(h), the school manager at Granhaug public school for children and youth with speech impairments, had already advised against enrolling him in an ordinary school.<sup>146</sup> Eng, who had published his own version of the Stanford–Binet scale that year, and was involved in debates on intelligence testing of children with hearing and speech difficulties, estimated the boy's IQ to be 80, although the father did not specify if the child had been tested.<sup>147</sup> The primary problem was that they resided in the neighboring Aker municipality, and Oslo schools typically did not accept children from

142 Sigurd Dahlstrøm, 'Johan Lofthus: Intelligensmåling', *Tidsskrift for Den Norske Legeforening*, 52.2 (1932), 76–77.

143 On Lofthus's work on the Sterilization Board, see Haave and Kyllingstad, chapter 4, this volume.

144 Johan Lofthus, Augusta Rasmussen and Bernhof Ribsskog, *Evnemålinger i Folkeskolens 1. Klasse* (Oslo: Gyldendal Norsk Forlga, 1937). On Ribsskog, see also Christian Ydesen, Brit Marie Hovland and Emma Vikstrøm, chapter 2, this volume.

145 BAO, Skoleetaten, Skolestyrets arkiv, 1913–31, Mappenr. 132, Letter from S.M. to Oslo Skolestyre, 09 March 1929.

146 Hans Eng was the brother of Norwegian psychologist and educator Helga Eng, see Paulsen, chapter 9, this volume.

147 Eng's version was printed as a special issue of the school magazine *Norsk Skoletidende*, see *Stanfordrevisjonen av Binet–Simons Intelligensprøver—Et Forsøk på Overføring til Norsk* (Særtrykk av Norsk Skoletidende nr. 24, 1929).

other areas. Nonetheless, the father requested an exception for special schools, highlighting his housing situation connected to his government work and relatively high salary. The letter was forwarded to the head teacher at Ullevålveien's school, who requested an examination of the child by Lofthus and mentioned available space at the school.<sup>148</sup> In August 1929, Lofthus concluded that the child was 'intellectually quite behind', with an IQ of 54, and 'could most possibly follow the first class at the special school'.<sup>149</sup> The Oslo School Board informed the father and the Aker school inspector that the child could attend the special school if the Aker School Board covered the expenses.<sup>150</sup> However, upon contacting the head teacher at Ullevålveien's school, the father learned that there was no longer space for his son.<sup>151</sup> He expressed willingness to cover the expenses himself to avoid delays if space became available the following year. Ultimately, the Aker School Board agreed to cover the expenses, and in March 1930, the father was informed that his son could attend the special school.<sup>152</sup>

The fate of the transfer application for the ten-year-old daughter of a construction worker residing in the working-class area of Aker municipality differed.<sup>153</sup> In August 1929, the Aker School Board inquired about moving the girl to Sarsgaten's school, as she had been given a place at Torshov school for the feeble-minded—a suggestion her parents opposed. They were more willing to consider the special school. The Oslo Special Schools Committee responded, requesting further details, including her IQ.<sup>154</sup> Lofthus examined her a month later, in September 1929, and penned a detailed report on her case.<sup>155</sup> Her mother accompanied her and provided valuable insights into their family situation and her daughter's health. The mother herself had repeated the first grade, but her father did well in school. The girl's

---

148 BAO, Skoleetaten, Skolestyrets arkiv, 1913–31, Mappenr. 132, Letter dated 09 March 1929.

149 BAO, Skoleetaten, Skolestyrets arkiv, 1913–31, Mappenr. 132, Letter dated 02 May 1929.

150 BAO, Skoleetaten, Skolestyrets arkiv, 1913–31, Mappenr. 132, Letter dated 21 September 1929.

151 BAO, Skoleetaten, Skolestyrets arkiv, 1913–31, Mappenr. 132, Letter dated 16 October 1929.

152 BAO, Skoleetaten, Skolestyrets arkiv, 1913–31, Mappenr. 132, Letter dated 07 March 1930.

153 BAO, Skoleetaten, Skolestyrets arkiv, 1913–31, Mappenr. 132, Letter dated 21 August 1929.

154 BAO, Skoleetaten, Skolestyrets arkiv, 1913–31, Mappenr. 132, Letter dated 19 September 1929.

155 BAO, Skoleetaten, Skolestyrets arkiv, 1913–31, Mappenr. 132, Letter dated 28 September 1929.

eldest sister had also repeated the first grade, while her two younger siblings were not yet attending school. Lofthus's report on the girl's family, particularly their history of repeating the first grade, may hint at his hereditarian views, although this is not explicitly stated. Despite overall good physical health, the girl herself had repeated the first grade. Her mother described her as possessing solid abilities and being reliable, capable of handling any task assigned to her. Lofthus concluded her IQ to be 63, suggesting it was not advisable for her to attend an ordinary school. However, he noted that her abilities aligned with those of younger students at the special school. In February 1930, the Oslo Special Schools Committee decided to reject the application, stating that 'with an IQ of 63, the child would fare better in a school for the feeble-minded', adding that both Sarsgaten's and Ullevålveien's schools were already overcrowded.<sup>156</sup> The final decision, indicating the child could not be accepted at Sarsgaten's school, was communicated back to Aker in March 1930.<sup>157</sup>

These two nearly parallel episodes exemplify core aspects of the Norwegian case that run through the whole chapter. While a psychiatric evaluation resulting in an IQ score was essential for evaluating a transfer request, it did not solely determine the outcome. The final decision depended on the school headmaster's broader resource considerations, the willingness of the relevant municipalities to cover the expenses, as well as on the parents' social status and their ability to advocate for their children. Although discussions both in Norway and internationally revolved around scientific authority and how test standardization could increase the trustworthiness of intelligence testing and secure a uniform process, these two incidents reveal a broader picture beyond standards. Medical versus educational insight, economic factors and administrative or geographical jurisdiction were most likely as important in such day-to-day transactions. Geographical boundaries on a local scale coincided with barriers to social movement, and thus exemplify the area's historical socio-economic divide. Parents' perceptions of the two types of schools, even if the distinction between special schools and schools for

---

156 BAO, Skoleetaten, Skolestyrets arkiv, 1913–31, Mappenr. 132, Letter dated 17 February 1930.

157 BAO, Skoleetaten, Skolestyrets arkiv, 1913–31, Mappenr. 132, Letter dated 1 March 1930.

the feeble-minded was blurred, made decisions more or less contentious independently of the alluring power of an IQ score.

The process of testing and its standardization merged school and educational problems with medical concerns, while practical and methodological issues folded into multiple societal, economic, professional, institutional, and personal dependencies. Standardization gave tests and testers the veneer of scientific work and the authority to make judgements, even when the resulting IQ score was not the sole criterion for decision making and the test was far from a perfect tool. The process of adapting the test acted as an important rhetorical and pragmatic argument for the diverse group of people brought together around the test in Norway. Thus, intelligence testing emerged as the 'boundary object' that held together a heterogeneous movement of educators, physicians, administrators, and parents.<sup>158</sup>

Even though earlier literature has rightly emphasized the leading role of psychiatrists in the Norwegian context, this study documents how standardization and testing depended on more intricate dynamics than professional sympathies or rivalries.<sup>159</sup> As psychiatry was still young, and psychology in its infancy, they both relied on the support of teachers to advance the agenda of testing. While the medicalization of education rightly points to the power of physicians, historians have shown how the rapprochement between medicine and schools was 'much less determined and much more contingent'.<sup>160</sup> In the Norwegian case, these networks included psychiatrists, pediatricians, psychologists, teachers, and politicians, test material and school testing spaces, as well as the discourses of hygiene, eugenics, and public health. The testing room was a crowded space that sought legitimation from exactly the multiplicity of voices it echoed.<sup>161</sup> In this space, discourses of care and control merged

158 For the idea of intelligence tests as boundary objects, see Andy Byford, 'The Mental Test as a Boundary Object in Early-20<sup>th</sup>-Century Russian Child Science', *History of the Human Sciences*, 27.4 (2014), 22–58.

159 Bjørn Hamre, Thom Axelsson, and Kari Ludvigsen, 'Psychiatry in the Sorting of Schoolchildren in Scandinavia 1920–1950: IQ Testing, Child Guidance Clinics, and Hospitalisation', *Paedagogica Historica*, 55.3 (2019), 391–415; Christian Ydesen, Kari Ludvigsen, and Christian Lundahl, 'Creating an Educational Testing Profession in Norway, Sweden and Denmark, 1910–1960', *European Educational Research Journal*, 12.1 (2013), 120–38.

160 Stephen Petrina, 'The Medicalization of Education: A Historiographic Synthesis', *History of Education Quarterly*, 46.4 (2006), 503–31 (p. 503).

161 For this idea I am indebted to my colleague, historian of medicine and museum



to create an amalgam of exclusionary inclusion. Here, standardization played its role in disciplining unruly testing alternatives and caring for a consistent scientific process, and at the same time provided an elusive guarantee of protection and modernization to society, and fairness or restriction of opportunity to individuals. However, from all the voices and purposes represented in this process, the most difficult to capture are those of the persons subjected to testing. A respectful reconsideration of the ethical guidelines regulating access to archives containing the stories of persons with disabilities might be a valuable corrective to this absence.<sup>162</sup>

Finally, this case suggests the centrality of movement to standardization. As a Norwegian story, this study is in line with historian Annette Mülberger's call for much-needed contextualization from a national perspective,<sup>163</sup> but at the same time allows the transnational to emerge. The chapter's focus on single individuals and their itineraries, especially as travel outside Norway constituted an important part of their education and further professional development, reveals how movement represented both negotiations of knowledge and accomplishment.<sup>164</sup> Traveling outside Norway brought international legitimation essential for embarking on the long process of standardization, while traveling within Norway was fundamental for ensuring collaboration and acceptance. Movement was also present as the result of standardization for schoolchildren transferred from one school to another. Similarly to human actors, intelligence tests crossed national borders, and test items were completely removed or moved around between age levels to achieve a version adapted to the local circumstances. However, the study proposes that as much as standardization depended on movement, it was equally the product of stasis—of staying long enough and persisting, remaining true to form, recognizable, bearing the signs of a revision.

---

curator Phil Loring.

162 Inger Marie Lid suggests a similar need for historical sources documenting possible harm to individuals in health-related situations, see 'A Work Farm for Deaf Men in Southern Norway', in *Islands of Extreme Exclusion*, pp. 140–41.

163 Annette Mülberger, 'The Need for Contextual Approaches to the History of Mental Testing', *History of Psychology*, 17.3 (2014), 177–86.

164 On how transnational aspects may be compatible with both a focus on a national frame and single individuals, see Soraya de Chadarevian, 'Commentary: Nationalism and Transnationalism in Anthropological Research', *Perspectives on Science*, 30.1 (2022), 194–98.

## Bibliography

- Aall, Anathon, 'Om en Ny Metode for at Maale Skole Dygtigheten hos', *Skolebladet* 18.2 (1915).
- 'Aandssvakhet Hyppigst på landsbygden både i England og i Norge', *Arbeiderbladet*, 45, 16 October 1930.
- Anchersen, Per, 'Ragnar Vogt 1870–1943', in *Psykologi og Psykologar i Norge*, ed. by Åse Gruda Skard (Oslo: Universitetsforlaget, 1959).
- Andresen, Astri and Kari Tove Elvbakken, 'From Poor Law Society to the Welfare State: School Meals in Norway 1890s–1950s', *Journal of Epidemiology and Community Health* (1979), 61.5 (2007).
- Axelsson, Thom, 'De Svåruppföstrade Barnen: Skolpsykiatrins Framväxt Och Etablering i Sverige 1910–1955', *Scandia*, 86.2 (2020).
- Barow, Thomas, "'Excluded from the Outside World": Planning and Discarding Island Institutions for "Degenerated Feeble-Minded (Asocial Imbeciles)" in 1920s Sweden', in *Islands of Extreme Exclusion: Studies on Global Practices of Isolation, Punishment, and Education of the Unwanted*, ed. by Bjørn Hamre and Lisa Villadsen (Leiden: Brill, 2024).
- 'Beretning om Nordisk Særskoleforbunds virksomhet i Norge', *Skolebladet* 24.36 (1921).
- Bergstrom-Katz, Sasha, 'On Intelligence Tests: Psychological Objects and Their Subjects' (Unpublished doctoral thesis, Birkbeck, University of London, 2023).
- Birkemo, Asbjørn, *Pedagogisk-Psykologisk Rådgiving* (Oslo: Universitet i Oslo, 2009).
- Byford, Andy, 'The Mental Test as a Boundary Object in Early-20<sup>th</sup>-Century Russian Child Science', *History of the Human Sciences*, 27.4 (2014).
- Carson, John, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750–1940* (Princeton: Princeton University Press, 2007).
- Clever, Iris, 'Miriam Tildesley and the Anthropological Politics of Standardizing Racial Measurements', *Perspectives on Science* 30.1 (2022).
- Dahlstrøm, Sigurd, *Studier over Unge Fanger. 1. Forsøk med Binet–Simons Intelligensprøver* (Kristiania: Aschehoug, 1922).
- Dahlstrøm, Sigurd, 'Arv og Miljø', *Tidsskrift for Den Norske Legeforening*, 49.12 (1929).
- Dahlstørn, Sigurd, 'Johan Lofthus: Intelligensmåling', *Tidsskrift for Den Norske Legeforening*, 52.2 (1932).

- de Chadarevian, Soraya, 'Commentary: Nationalism and Transnationalism in Anthropological Research', *Perspectives on Science*, 30.1 (2022).
- Det Kongelige Frederiks Universitets Årsberetning, 1ste Juli 1922–30te Juni 1923 (Oslo: A. W. Brøggers, 1923).
- Det Kongelige Frederiks Universitets Årsberetning, 1ste Juli 1923–30te Juni 1924 (Oslo: A. W. Brøggers, 1926).
- 'Det Store Fylkesmøte i Elverum for Fysisk Opfostring og Skolehygiene', *Østlendingen*, 83, 07 November 1924.
- 'Direktiver og Instrumentarium til Intelligensmaaling', *Tidsskrift for Den Norske Legeforening*, 11 (1925).
- Eklund Koza, Julia, "*Destined to Fail*": *Carl Seashore's World of Eugenics, Psychology, Education, and Music* (Ann Arbor: University of Michigan Press, 2021).
- 'En Mann som Vil Måle Vår Intelligens', *Arbeiderbladet*, 265, 29 September 1931.
- Eng, Hans, *Stanfordrevisjonen av Binet–Simons Intelligensprøver–Et Forsøk på Overføring til Norsk* (Særtrykk av Norsk Skoletidende nr. 24, 1929).
- Froestad, Jan and Bodil Ravneberg, 'Education Policy, the Norwegian Unitary School and the Social Construction of Disability', *Scandinavian Journal of History*, 31.2 (2006).
- Frøyland Nielsen, 'Julie Monrad Jacobsen Død', *Vor Skole*, 49.1–51/52 (1963).
- Gjermoe, Eilif, 'Særskolestatistikk', *Hjälpskolan Værneskolen Særskolen* 6.1 (1927).
- Gruda Skard, Åse, 'Martin L. Reymert 1883–1953', in *Psykologi og Psykologar i Norge*, ed. by Åse Gruda Skard (Oslo: Universitetsforlaget, 1959).
- Haave, Per, 'Psykiatrien og De Vanskeligste Skolehjemsbarna: Tanke og Handling Mellom To Kriser' (forthcoming).
- Hagemann, Gro, *Skolefolk: Lærernes Historie i Norge* (Oslo: Ad Notam Gyldendal, 1992).
- Hamre, Bjørn, Axelsson, Thom and Kari Ludvigsen, 'Psychiatry in the Sorting of Schoolchildren in Scandinavia 1920–1950: IQ Testing, Child Guidance Clinics, and Hospitalisation', *Paedagogica Historica*, 55.3 (2019).
- Havik, Håkon, *Mål Og Mening: Idéhistorisk Blikk På Johan Lofthus' Intelligensmålinger i Norge 1918–1939* (Oslo: Universitetet i Oslo, 2006).
- 'Hva Er Intelligens?', *Aftenposten*, 488, 28 September 1931.
- 'Intelligens i Universitetets Gamle Festsal', *Dagbladet*, 226, 28 September 1931.
- 'Intelligens og Evner i Skolen', *Dagbladet*, 45, 22 February 1930.
- 'Intelligensprøver til Bruk Særlig for Asyl og Rettspsykiatriske Sakkyndige', *Tidsskrift for Den Norske Legeforening*, 17–18 (1924).

- Kirkebæk, Birgit, 'Sexuality as Disability: The Women on Sprogø and Danish Society', *Scandinavian Journal of Disability Research*, 7.3–4 (2005).
- Kobbeestad, Egill, 'Fylkesmøte for Fysisk Opfostring og Skolehygiene', *Østlandet*, 282, 08 December 1924.
- Larse Bryne, Lina, "'Med Maaling Følger Viden': Intelligensmåling i den Norske Folkeskolen 1900–1945' (Unpublished master's thesis, University of Bergen, 2006).
- Lefkaditou, Angeliki, "'This Wonderful People': Darwin, the Victorians, and the Greeks', *Journal of Modern Greek Studies*, 36.1 (2018).
- Lefkaditou, Angeliki, 'Naturalizing the Nation: Physical Anthropology in Greece, 1880s–1950s' (Unpublished doctoral thesis, University of Leeds, 2021), [https://etheses.whiterose.ac.uk/id/eprint/30102/1/Lefkaditou\\_PhD\\_Naturalizing%20the%20Nation.pdf](https://etheses.whiterose.ac.uk/id/eprint/30102/1/Lefkaditou_PhD_Naturalizing%20the%20Nation.pdf)
- Lid, Inger Marie, 'A Work Farm for Deaf Men in Southern Norway', in *Islands of Extreme Exclusion: Studies on Global Practices of Isolation, Punishment, and Education of the Unwanted*, ed. by Bjørn Hamre and Lisa Villadsen (Leiden: Brill, 2024).
- Lofthus, Johan, 'Intelligensundersøkelser og Psykiatrisk Arbeide ved Kristiania Komm. Skoler', in *Nordisk Kongress for Fysisk Opfostring og Skolehygiene i Kristiania 28. septbr.–1. Oktober 1924*, ed. by Otto Grennes (Oslo: Eberh. B. Oppi forlag, 1925).
- Lofthus, Johan, 'Om Intelligensmåling og Særskoleundervisning', *Norsk Skoletidende*, 38 (1927).
- Lofthus, Johan, *Intelligensmåling* (Oslo: H. Aschehoug & Co., 1931).
- Lofthus, Johan, 'Skolepsykiaterens Arbeid', in *Oslo Særskole: Gjennom 50 År*, ed. by Signe Gran and Julie Monrad Jacobsen (Oslo: Thronsen & Co., 1943).
- Lofthus, Johan, Augusta Rasmussen and Bernhof Ribsskog, *Evnemålinger i Folkeskolens 1. Klasse* (Oslo: Gyldendal Norsk Forlga, 1937).
- Looft, Carl, *Kliniske og Ætiologiske Studier over Psykiske Udviklingsmangler hos Børn* (Bergen: John Grieg, 1897).
- Looft, Carl, 'Intelligensundersøkelser av skolebarn', *Medicinsk Revue*, 10 (1913).
- Looft, Carl, 'Sinkers og Aandsvake Barns Skoleforhold i Bergen', *Tidsskrift for Den Norske Legeforening*, 45.4 (1925).
- Lov om Folkeskolen for Kjøbstæderne: Stockholms Slot den 26de Juni 1889* (Kristiania: P. T. Mallings Boghandels Forlag, 1889).
- Lov om Folkeskolen paa Landet: Stockholms Slot den 26de Juni 1889* (Kristiania: P. T. Mallings Boghandels Forlag, 1889).
- Magnussen, Svein, Killingmo, Bjørn and Siri Erika Gullestad, *Klinikk og Laboratorium: Psykologi i Hundre År* (Oslo: Universitetsforlag, 2009).

- McGuire, Coreen, *Measuring Difference, Numbering Normal: Setting the Standards for Disability in the Interwar Period* (Manchester: Manchester University Press, 2020).
- Moljord, Gøril and Hilde Bondevik, 'Constructions of the In-/Educable: A Nordic Outlook on Changing Legal and Educational Statuses of Pupils with Disabilities', in *The Nordic Education Model in Context: Historical Developments and Current Renegotiations in Studies in Curriculum Theory*, ed. by Daniel Tröhler (New York: Routledge, 2023).
- Monrad Jacobsen, Julie, 'Særeskolen, Hvad Den Er, Hvad Den Vil, Og Hvordeles Den Arbeider', in *Elleve Nordiske Skolemøte i Kristiania 5–8 August 1920*, ed. by O. J. Hoversholm, Magnus Alfsen and G. E. Mæhlum (Kristiania: Fabritius, 1921), pp. 92–106.
- Monrad Jacobsen, Julie, 'Fra et Besøk ved Tyske Skoler', *Vor Skole*, 10.20 (1921).
- Monrad Jacobsen, Julie, *Særeskolen, Dens Maal og Midler* (Göteborg, 1922).
- Monrad Jacobsen, Julie, 'The Special School in Norway and Work Connected with It', *Studies in Mental Inefficiency*, 15.3(2) (1922), 25–31.
- Myers, Charles S., ed., *VIIth International Congress of Psychology Held at Oxford from July 26 to August 2, 1923, under the Presidency of Charles S. Myers, Proceedings and Papers*, edited by Charles S. Myers (Cambridge: Cambridge University Press, 1924).
- Mülberger, Annette, ed., *Mental Testing after 1905: Uses in Different Local Contexts*, special issue of *History of Psychology*, 17.3 (2014).
- Mülberger, Annette, 'The Need for Contextual Approaches to the History of Mental Testing', *History of Psychology*, 17.3 (2014).
- Mülberger, Annette, 'Biographies of a Scientific Subject: The Intelligence Test', in *Oxford Research Encyclopedias: Psychology* (Oxford: Oxford University Press, 2020), <https://doi.org/10.1093/acrefore/9780190236557.013.694>
- 'Nordisk Særskoleforbund', in *Oslo Særskole: Gjennom 50 År*, ed. by Signe Gran and Julie Monrad Jacobsen (Oslo: Thronsen & co, 1943).
- 'Norsk Psykiatrisk Forening', *Tidsskrift for Den Norske Legeforening*, 17–18 (1924).
- Petrina, Stephen, 'The Medicalization of Education: A Historiographic Synthesis', *History of Education Quarterly*, 46.4 (2006).
- 'Psykologi og Intelligensprøver i Norge', *Aftenposten*, 619, 25 October 1923.
- Ravneberg, Bodil, 'Grensebarna i Skolen. Utskillingen av "Sinkene" fra Normal- og Åndssvakekategoriene Tidlig på 1900-tallet', in *Funksjonshemming, Politikk og Samfunn*, ed. by Jan Froestad, Per Solvang and Mårten Söder (Oslo: Gyldendal akademisk, 2000).
- Reymert, Martin L., *Stanford Revisjon av Binet–Simons Intelligensprøver: Forsøk paa Tilpasning til Bruk for Norsk Standardisering* (Oslo: Arthur Nelson, 1924).

- Reymert, Martin L. and John Frings, 'Children's Intelligence in Relation to Occupation of Father', *American Journal of Sociology*, 41.3 (1935).
- Rust, Val D., *The Democratic Tradition and the Evolution of Schooling in Norway* (New York: Greenwood Press, 1989).
- Saarheim, Hjørdis, 'Differentiering i Folkeskolen', *Norsk Skoletidende*, 10 (1930).
- Schiøtz, Carl, 'Helsen nr. 1, undervisningen nr. 2', in *Nordisk Kongress for Fysisk Opfostring og Skolehygiene i Kristiania 28. septbr.–1. Oktober 1924*, ed. by Otto Grennes (Eberh. B. Oppi forlag: Oslo, 1925).
- Simonsen, Eva, *Vitenskap og Profesjonskamp: Opplæring av Døve og Åndssvake i Norge 1881–1963*, 2 (Oslo: Unipub forlag, 2000).
- Studentene fra 1907: Biografiske Oplysninger Samlet til 25-Års-Jubileet 1932* (Oslo: Grøndhal & Søn, 1932).
- Sunde, Anton, 'Sigurd Dahlstrøm in memoriam', *Tidsskrift for Den Norske Lægeforening*, 53.9 (1933), 453–55.
- Sørensen, Marta, 'Intelligensprøver med Barn', *Vor Skole*, 11.37 (1922).
- Terman, Lewis M., *The Measurement of Intelligence: An Explanation of and a Complete Guide for the Use of the Stanford Revision and Extension of the Binet–Simon Intelligence Scale* (Boston: Houghton Mifflin Co., 1916).
- Terman, Lewis M., *Record Booklet for the Stanford Revision of the Binet–Simon Tests as Described in Terman's The Measurement of Intelligence* (Boston: Houghton–Mifflin, 1916).
- 'Til Styret for Forskningsfondet av 1919', *Skolebladet* 24.44 (1921).
- Thuen, Harald, *Den Norske Skolen: Utdanningssystemets Historie* (Oslo: Abstrakt forlag, 2017).
- Thompson, Matthew, *The Problem of Mental Deficiency: Eugenics, Democracy, and Social Policy in Britain, c.1870–1959* (Oxford: Clarendon, 1998).
- Vogt, Ragnar, 'Standardisering af Intelligensprøver for Norske Barn', *Aftenposten*, 543, 12 November 1921.
- Vogt, Ragnar, 'Maaling av Intelligens', *Morgenbladet*, 64, 26 February 1921.
- Ydesen, Christian, Kari Ludvigsen and Christian Lundahl, 'Creating an Educational Testing Profession in Norway, Sweden and Denmark, 1910–1960', *European Educational Research Journal*, 12.1 (2013).
- Zenderland, Leila, *Measuring Minds: Henry Herbert Goddard and the Origins of American Intelligence Testing* (Cambridge, UK: Cambridge University Press, 1998).



## 2. The Scandinavian Space of IQ Testing: Between Normal and Special Education, 1918–1940

*Christian Ydesen, Brit Marie Hovland, and  
Emma Vikström*

---

The evolution of IQ testing technology and its impact on conceptions of human intelligence have consistently contained a prominent transnational dimension.<sup>1</sup> Beginning with the introduction of the standardized Binet–Simon test in Paris in 1908, different IQ tests soon proliferated across Europe, the US, and beyond, including in Scandinavia, where they were translated and adapted into national and/or local educational practices. As argued by Martin Lawn in his analysis of testing and statistics in the twentieth century: ‘Testing and its devices flowed across borders into other systems of education, and the publication of test results and standards of work “travelled” across scientific communities’.<sup>2</sup>

This chapter investigates the nature and significance of inspirations, experimental endeavours, and exchanges among key actors and knowledge brokers involved in IQ testing and educational reform in Norway, Sweden, and Denmark during the interwar period. These

---

1 Annette Mülberger, ‘The Need for Contextual Approaches to the History of Mental Testing’, *History of Psychology*, 17 (2014), 177–86; Christian Ydesen, *The Rise of High-Stakes Educational Testing in Denmark, 1920–1970* (Doctoral thesis, Frankfurt a.M: Peter Lang Verlag, 2011).

2 Martin Lawn, ‘Voyages of Measurement in Education in the Twentieth Century: Experts, Tools and Centres’, *European Educational Research Journal*, 12 (2013), 108–119 (p. 109).



neighboring Scandinavian countries shared a robust level of exchange, not least due to their linguistic affinity and mutual perception of each other as reference societies. Through this analysis, the chapter elucidates the emergence and institutionalization of IQ testing within the intersecting domains of regular education and the changing provisions around special schooling in these three nation-states, with some additional emphasis given to Norway. Furthermore, the study provides insights into the IQ-related selection and sorting into special education from the regular school at the primary and lower secondary levels [*folkeskole*], as a necessary condition for the early foundations and political vision of an unitary school [*enhetsskole*] in the budding Nordic welfare states.<sup>3</sup> In other words, the Scandinavian one-school-for-all system had, as its implicit and symbiotic counterpart, the special schools for children who did not fit into regular education. In this sense, the chapter contributes to our understanding of how IQ testing and the concept of “intelligence” came to function as a technology that shaped policies and practices related to normality and deviance—and, in turn, influenced the development of unitary schooling in Scandinavia and, more broadly, the Nordic model of education.<sup>4</sup>

The chapter examines the inter-Scandinavian special education journal *Hjælpskolan/Værneskolen/Særskolen*,<sup>5</sup> first published in 1923,

- 3 Daniel Tröhler, ‘Introduction: The Nordic education model’, in *The Nordic Education Model in Context*, ed. by Daniel Tröhler, Bernadette Hörmann, Sverre Tveit and Inga Bostad (New York: Routledge, 2022), pp. 1–11. The concept of the unitary school, often referred to as “comprehensive school” within educational sciences, aims ideally to provide equal opportunities for all individuals, promoting a meritocratic system irrespective of social background. Consequently, this also involves identifying and excluding those deemed “uneducable”, as they are perceived as impediments to the objectives of the unitary school. It is important to note that the term *folkeskole* [primary and lower secondary school] was used to refer to public primary schools in all three countries during the early twentieth century, but that the unitary aspect was realized in different time periods. In Norway, the term *folkeskole* was formally introduced in 1889, replacing the former *almueskole* [the common school], and a similar change occurred in Denmark with the 1899 education act. In Sweden, the unitary school was not formally introduced until the early 1960s.
- 4 It is common for “Scandinavia” and the “Nordic” region, which also includes Finland, Iceland, and the Faroe Islands, to be used interchangeably in both historical and contemporary discussions. Notably, the historical figures and knowledge brokers discussed in this chapter predominantly used the term “Nordic” in their writings. To avoid confusion, we will adhere to the terminology used in the sources while maintaining our analytical focus on Norway, Sweden, and Denmark as denoted by the concept “Scandinavia”.
- 5 The English translations of these terms are “The Help School,” “The Care School,” and

alongside archival materials, newspaper clippings, and policy documents. We found that Scandinavian actors shared a common recognition of the educational challenges involved in establishing a unitary school system. This enabled them to draw on practices and knowledge from neighbouring countries as reference points for their own educational and scientific endeavours, with significant implications for the development and use of intelligence assessment across the three countries. Moreover, these actors collectively grappled with the dilemma of whether and how to use tests in designing a reliable, science-based system for determining inclusion and exclusion within an educational project closely tied to broader goals of social reform and the formation of the Nordic welfare states.<sup>6</sup>

## Inclusion and Exclusion in Nordic Unitary Schooling

In contemporary comparative and international education studies, the so-called Nordic education model often serves as a reference point.<sup>7</sup> Historically, the Nordic education systems from the nineteenth century grew from a political vision of a free and accessible unitary school for all, and a social reformatory vision of a connecting line through the

---

“The Special School.” The variation in terminology across the three countries reflects different cultural and historical connotations. “Help” suggested that a child required special assistance due to specific needs; “Care” implied that a child needed to be shielded from society; while “Special” conveyed the need for alternative provision. At the fifth meeting of the association in Norway in 1934, the question of the most appropriate name for the special school among Nordic colleagues was discussed. Referring to practices in Sweden and Finland, *Hjælpeskolen* [the help school] was advocated as a common Scandinavian term. The main argument focused on the potential stigmatizing effects of “remedial education”. However, the suggestion was not taken up, as participants argued that the terminology was less important than the cultural effort of enlightening teachers and parents about the positive potential of special education. Elisabeth Faye-Hansen (see below) further noted that the term “help classes” already acquired a negative reputation in Oslo. Kristian Haunso, ‘Hvad skal vi kalde skolen for de svagt begavede?’ *Hjælpskolen/Værneskolen/Særskolen*, 12 (1935), pp. 8–14.

6 Pauli Kettunen, ‘The Transnational Construction of National Challenges: The Ambiguous Nordic Model of Welfare and Competitiveness’, in *Beyond Welfare State Models: Transnational Historical Perspectives on Social Policy*, ed. by Pauli Kettunen and Klaus Petersen (Cheltenham: Edward Elgar, 2011), pp. 16–40.

7 John Benedicto Krejsler, ‘The Nordic dimension as a metaspace for educational research’, *Nordic Studies in Education*, 43 (2023), 8–24; Christian Ydesen, ‘The Shadowy Existence of the Nordic Dimension in Education’, *Nordic Studies in Education*, 43 (2023), 260–265.

educational system for the entire population.<sup>8</sup> At its origins, this vision for a unitary school was firmly rooted in broader political concerns about nation-building.<sup>9</sup> However, as the number of enrolled pupils in unitary schooling increased from 1900 onwards, a new demand for differentiation became evident. As such, special education and the handling of so-called “backward” children became a key concern in Nordic educational debates. The special school was seen as ‘the extreme outpost’ of the primary and lower secondary school, taking care of the so-called ‘residual group’, i.e., children that fit neither into normal schools nor mental care institutions.<sup>10</sup> Illustrating the intricate connections between pedagogy, psychology, and psychiatry, and drawing upon the influence of the international child guidance movement, one can observe efforts to establish pedagogical-psychological counseling offices and societies in Norway, Denmark, and Sweden, albeit at varying rates, from the interwar period onwards.<sup>11</sup>

In the interwar years, the dialectic between inclusion and exclusion hence implied a need for practices that would allow more objective, scientific identification of normality and deviance. In this quest, notions on empirical measurement technologies of intelligence, as well as the development and use of IQ tests, were a *sine qua non* for educators, psychologists, psychiatrists, and medical doctors engaged in the work to develop a well-functioning unitary education and welfare system

8 Harald Thuen, *Den norske skolen: Utdanningssystemets historie* (Oslo: Abstrakt forlag, 2017).

9 Brit Marie Hovland, ‘Mapping the Nation—Fabricating Citizens from Patriotism Towards Nordicism?’, in *The Nordic Education Model in Context*, ed. by Daniel Tröhler, Bernadette Hörmann, Sverre Tveit and Inga Bostad (New York: Routledge, 2022), pp. 15–35.

10 Eva Simonsen, “‘Folkeskolens Ytterste Forpost’—En Annen Historie Om Den Norske Folkeskolen 1945–1970’, *Uddannelseshistorie*, 26 (2012), 118–44; Gøril Moljord and Hilde Bondevik, ‘Constructions of the In-/Educable’ in *The Nordic Education Model in Context*, ed. by Daniel Tröhler, Bernadette Hörmann, Sverre Tveit and Inga Bostad (New York: Routledge, 2022), pp. 56–74; Bjørn Hamre, ‘Den yderste halespids: Bekymring som dispositiv i skolens problemforståelser og forskelssætninger’, *Uddannelseshistorie*, 46 (2012), 70–94; Bjørn Hamre, Thom Axelsson, and Kari Ludvigsen, ‘Psychiatry in the Sorting of Schoolchildren in Scandinavia 1920–1950: IQ Testing, Child Guidance Clinics, and Hospitalisation’, *Paedagogica Historica*, 55 (2019), 391–415.

11 Ydesen, *The Rise of High-Stakes Educational Testing in Denmark*, p. 79f.; Christian Lundahl, ‘Making testers out of teachers: The work of a Swedish State Research Institute 1946–1956’, *History of Education* 48 (2019), 646–63; Åse Gruda Skard, ‘Arbeidet ved en mentalhygienisk skuleklinikk’, *Norsk pedagogisk årbok* 7 (1934–35), 75–80.

able to cater for all types of educational needs. The architects of public education across Scandinavia were involved in drawing inspiration, conducting experimental work, and exchanging ideas with their counterparts in foreign education systems, especially with their Nordic neighbors.<sup>12</sup>

From the late nineteenth century, educational exchanges became institutionalized through inter-Nordic associations, meetings, courses, and journals. A clear example of this institutionalization, particularly in the establishment of an “outpost” for the primary education system, can be found in *Decorah-Posten*—a Norwegian-language newspaper published in Iowa, USA—which featured the following article in 1928:

Nordic Special School Meeting. Last Sunday, at Hindsgavl Castle near Middelfart [Denmark], more than 100 educators from Norway, Sweden, Finland and Denmark gathered at a meeting that lasted 4 days to discuss what can be done for backward children and young people.<sup>13</sup>

The mere existence of this newspaper article indicates that special education in the Nordic region garnered attention far beyond its borders, demonstrating a broad interest in the subject. Of potentially greater significance, however, is its content, which underscores the transnational intermeshing of intelligence and IQ testing in Scandinavia. This intermeshing had already been occurring before its formal institutionalisation in 1920 with the establishment of a committee for the Nordic Special School Association [*Nordiskt Hjälpsskoleförbund/Værneskoleforbund/Særskoleforbund*]. During its inaugural congress in Dalsæter, Norway, in the summer of 1922, the decision was made to launch a joint Nordic journal titled *Hjälpsskolan/Værneskolen/Særskolen*, which published its first volume the following year.

12 Kari Ludvigsen, Christian Lundahl and Christian Ydesen, ‘Creating an educational testing profession—the emergence and impact of the Scandinavian testing community, 1920–1960’, *European Educational Research Journal*, 12 (2013), 120–138.

13 ‘Nordisk Værneskolemøde’, *Decorah-Posten*, 21 August 1928, p. 3. All translations are made by the authors unless stated otherwise. The meeting was organized by the Nordic Association of Special Schools, whose president was Headmaster Jørgen Olsen, Copenhagen. Among the speakers of the meeting can be mentioned: Dr. med. Johan Lofthus, Oslo; special-education teacher Alice Descoeudres, Geneva; teacher’s college teacher Ruth Claeson, Stockholm; principal Hermanda Stilling, Copenhagen; head teacher M. Bording, Vejle; mag. art. Sofie Riffbjerg, Copenhagen; principal Ulf Hässelberg, Gothenburg; superintendent Signe Elgenström, Stockholm; and school manager Marie Pedersen, Trondheim.

## Sources, Methodology, and Theoretical Underpinnings

Using the methodological toolbox of historical source criticism, our analysis explores the appearance and agency of key actors and knowledge brokers from Norway, Sweden, and Denmark, their activities presenting and discussing notions about intelligence, and the developments and institutionalization of IQ test practices. As source material, we have collected all issues of the educational journal *Hjälpskolan/Værneskolen/Særskolen* (1923–1940), the in- and outgoing correspondence in archives of the Danish Society for Experimental Pedagogy [*Foreningen for Experimental Pædagogik*] and the Danish Education Psychology Study Commission [*Udvalget for Skolepsykologiske Undersøgelser*], as well as Norwegian newspaper and educational journal articles from the period.<sup>14</sup>

Inspired by prosopographic methodology—investigating collective biographies through networks and organizations—we emphasize the potential in situating agents, events, and practices spatially. This opens up perspectives for investigating the distribution of knowledge, and consequently the re-interpretation and rearticulation of knowledge, ideas, and practices, both within and beyond institutions.<sup>15</sup> More specifically, this approach provides insights into the complex network that spanned emerging borders of academic and professional disciplines across the three Scandinavian countries and beyond. It gives us indications about the discursive and practical negotiations of IQ testing and intelligence within the field of education, and how this new technology influenced school politics, school system reforms, and practices of sorting children into different school streams according to their “*Bildsamkeit*”, or educability, as determined by IQ testing.

Our analysis of the spaces of exchange, particularly regarding the development of ideas about IQ testing and conceptions of intelligence, is inspired by Bruno Latour’s concept of the ‘center of calculation’. This concept also refers to spatial venues where knowledge production relies

14 Our search terms in the database of the Norwegian National Library included the names of the actors associated with the Nordic Special School Association, as well as concepts such as “*intelligens*”, “*intelligenstest*”, “*særskole*”, “*værneskole*”, “*hjælpekole*” (in various Scandinavian spellings).

15 Verboven Koenraad, Myriam Carlier and Jan Dumolyn, ‘A short manual to the art of prosopography’, in *Prosopography Approaches and Applications. A Handbook* (Oxford: Unit for Prosopographical Research, Linacre College, 2007), pp. 35–69.

on the accumulation of resources through the circulatory movements of actors between different locations.<sup>16</sup> Latour describes how centers of calculation project power by systematically assembling and expanding a material network of actors.<sup>17</sup> In this context, “intelligence” is sustained, reproduced, and developed by a community of transnational knowledge brokers—key intermediaries that facilitate the circulation of knowledge across borders and contexts.<sup>18</sup> The functioning of this network shapes the forms and uses of IQ tests and notions of intelligence at Nordic Special School Meetings and in the analyzed journal. Both sites reflect the specific national contexts of emerging unitary schooling frameworks in Norway, Sweden, and Denmark, all grappling with similar societal and educational challenges.

## The Inter-Scandinavian Context of IQ Testing and “Intelligence”

To comprehend the context of Scandinavian IQ testing and the concept of intelligence in education, it is essential to recognize the inherent and often implicit Nordic dimension. A sense of Nordic community appeared to shape the frame of reference for many educators, psychologists, psychiatrists, and medical doctors involved in IQ testing and intelligence, leading them to view neighboring countries as reference societies and their counterparts in these countries as collaborators.<sup>19</sup> This sense of

16 Heike Jöns, ‘Centre of calculation’, in *The SAGE Handbook of Geographical Knowledge*, ed. by John Agnew and David N. Livingstone (London: SAGE Publications, 2011), pp. 158–170.

17 Jason Beech and Alejandro Artopoulos, ‘Interpreting the circulation of educational discourse across space: Searching for new vocabularies’, *Globalisation, Societies and Education*, 14 (2016), 251–271; Thomas S. Popkewitz, ‘Comparative reasoning, fabrication, and International Education Assessments: Desires about nations, society, and populations’, *International Journal of Educational Research*, 112 (2022).

18 See also Matthew S. Weber and Itzhak Yanovitzky, ‘Knowledge Brokers, Networks, and the Policymaking Process’, in *Networks, Knowledge Brokers and the Public Policymaking Process*, ed. by Matthew S. Weber and Itzhak Yanovitzky (New York: Palgrave, 2021), pp. 1–25.

19 It should be duly mentioned that the Nordic and Scandinavian spaces of special education and IQ testing were part of a larger European and global space. As such most of the actors and knowledge brokers were well-versed in international developments and many of them had studied in Germany, which shaped their outlooks and professional trajectories. Simultaneously, there was also an interest in the Nordic region from abroad.

community within the Nordic educational sphere can be traced back to national teacher associations formed during the mid-nineteenth-century nation-building processes, with the first Nordic school meeting occurring in Gothenburg in 1870.<sup>20</sup> From the late 1860s onward, the organization of Nordic meetings and periodicals also emerged in the field of special education. By the interwar years, a well-established tradition of Scandinavian exchanges and institutionalization in special education provided fertile ground for the formation of the new Nordic special school association, their journal, as well as the exchange and use of IQ tests.

### Reforms Concerning Special Education in the Three Scandinavian Countries

The expansion of education and the development of unitary schools in all three Scandinavian countries led to a more heterogeneous pupil population, prompting calls for differentiation based on pupils' abilities and educational needs. In Norway, the foundation for systematic special education was laid with the 1881 adoption of the School Act on the Education of Abnormal Children [*Lov om abnorme børns undervisning*], a precursor to more comprehensive legislation. The 1889 People's School Acts [*Folkeskolelovene*] further institutionalized special education, establishing special classes as a response to the emerging concept of a "normal" school. The very term *folkeskole*, which literally translates as 'the people's school', reflected both socio-economic and national-political arguments for a new, unified national education system, while simultaneously accommodating special education as an excretion from this new normality:

For Children, whom the School Board does not deem able to participate in the regular school, when the Municipal Council has given its consent, special school education shall be arranged; the expenses, hereby incurred to the Municipality, can be fully or partial[ly] reimbursed by the relevant Providers.<sup>21</sup>

20 Jón Torfi Jónasson and Valgerður S. Bjarnadóttir, 'The Nordic interactive and comparative spaces within the arena of education', *Nordic Studies in Education*, 43 (2023), 25–42; Brit Marie Hovland, *For fedreland og broderband: Allmugeskulelærarane si møteverksemd i det nasjonale gjennombrøtet 1850–1870* (Volda: Høgskolen i Volda, 1999).

21 People's School Act for Cities (1889), Chapter 1, § 6. Introducing one school for all,

The institutionalization of this new unitary school ‘for all’ inevitably raised the need to define the boundaries of normality and its limitations. Together, the 1881 Abnormal School Act, the 1889 People’s School Act, and, later on the 1896 law on child welfare boards and institutions [*Vergerådsloven*], established a framework of segregation embedded within a discourse of normality. The ‘normal’ school could exclude ‘those with significant learning difficulties or poor health who struggled to keep up’.<sup>22</sup> This system evolved rapidly. The Unitary School Committee of 1911, and major school commissions in the 1920s, engaged in extensive debate over the role and organization of special education. The 1936 School Act and the 1939 Normal Plan reaffirmed the vision of a one-school-for-all system and promoted the prevailing idea of a normal school intended for the normal pupil. It included special classes within regular schools for students with IQs between 70 and 90, municipal special schools for those in the same range, and national special schools for pupils with IQs below 70. Taken together, these developments reveal that alongside the Norwegian ‘normal’ school system, a parallel system of special education evolved, ultimately leading to the Special School Acts of 1951 [*Lov om spesialskolene*], which established the Directorate of Special Schools under the leadership of Marie Pedersen (1893–1990).<sup>23</sup>

In Denmark, the origins of public special education date back to 1891, when the municipality of Frederiksberg created a special class for children classified as “mentally retarded” or disabled. Between 1919 and 1923, the Great School Commission recommended the establishment of special classes wherever feasible. However, due to the highly decentralized Danish education system, where local authorities held significant power over public schooling, it was not until 1937 that national legislation was enacted to regulate special education as part of the overall education system. This legislation stipulated that children

---

and secular school subjects.

22 People School Act 1889, Chapter 1, §4. See Thuen, *Den norske skolen*, p. 112, pp. 122–123; Harald Thuen, *Om barnet: Oppdragelse, opplæring og omsorg gjennom historien* (Oslo: Abstrakt forlag, 2008), pp. 134–141.

23 Marie Pedersen translated the Dearborn group test into Norwegian, which later served as the basis for the Danish versions, and she became director of the Special School in Oslo in 1939. Her background, pivotal contributions, and the group test’s significant influence on testing regimes and school politics during the interwar period and beyond fall outside the scope of this chapter, which places its emphasis on individual IQ tests and testing.



who could not keep pace with regular instruction should be transferred to specialized classes, provided local conditions allowed.<sup>24</sup> Children with IQs between 90 and 110 attended normal schools, those with IQs between 70 and 90 were placed in special schools, and children with IQs below 70 were assigned to mental care institutions.<sup>25</sup>

In Sweden, the development of special schools and classes can be partly attributed to the expansion of primary education in the early 1900s. As the primary school system expanded, the need arose to provide education for children who did not fit into neither normal schools nor mental care institutions for the feeble-minded [*sinneslöanstalter*].<sup>26</sup> Special schools and classes were created to serve children deemed backward and unable to meet the demands of regular education. A 1936 Swedish policy recommendation stated that children with IQs between 70 and 80 should be transferred to special classes. Borderline cases also required consideration of psychological, social, and physical factors.<sup>27</sup>

The differentiation into various forms of special education was justified by the notion of varying levels of ability, an increasingly influential argument within the emerging one-school-for-all model. In all three countries, many political advocates of the unitary school saw few issues with the exclusion of pupils deemed backward. Conversely, they argued that these pupils would receive a more suitable education in special schools, tailored to their individual needs.<sup>28</sup> Others insisted that the IQ tests were inadequate and incomplete. The practice of sorting children remained controversial. The controversy reveals disciplinary boundaries between professional groups; pedagogues and teachers against more psychology-oriented pedagogues, medical doctors, and psychiatrists. A key justification for differentiation based on giftedness was the empiricist and positivist belief that students

24 Ydesen, *The Rise of High-Stakes Educational Testing in Denmark*; Frederik C. Kaalund-Jørgensen, 'Hvad gør vi for de Børn, der ikke kan følge Folkeskolens almindelige Undervisning?', *Hjælpskolan/Værneskolen/Særskolen*, 20 (1942), 87–95.

25 Ydesen, *The Rise of High-Stakes Educational Testing in Denmark*, p. 58.

26 Thom Axelsson, *Rätt elev i rätt klass: Skola, begåvning och styrning 1910–1950* (Doctoral thesis, Linköping: Linköping's University, 2007), p. 74.

27 Statens offentliga utredningar, 1936:31, *Betänkande med utredning och förslag angående åtgärder för särskild undervisning och utbildning av psykiskt efterblivna barn i barn- och ungdomsåren* (Stockholm: Ivar Haeggströms boktryckeri A.B, 1936), p. 100.

28 For the Swedish case, see e.g. Axelsson, *Rätt elev i rätt klass*, p. 65.

would be assessed using ostensibly fair, objective, and scientific criteria, rather than socio-economic background, or teachers' unscientific professional judgment. IQ tests were underscored as a more neutral and objective standard and progress.<sup>29</sup> However, most educators and many psychologists insisted on the teachers' professional opinion and their personal knowledge of the child and family. Instead of personal IQ tests to delineate between normal and abnormal, group tests performed by the teachers in their class were introduced as a mapping instrument for early intervention.<sup>30</sup>

### The Nordic Special School Association: Some Prosopographic Characteristics

The initial idea for establishing a formal Nordic Special School Association was aired in 1916 at the Norwegian Schoolmistress Association meeting where participants from all three Scandinavian countries took part.<sup>31</sup> The committee of the new budding association was institutionalized in 1920. Organized in a tri-annual meeting cycle from 1922 onwards, the Nordic Special School Association—and its journal from 1923—bears all the characteristics of a significant transnational arena for exchange, a center of calculation.<sup>32</sup> It was a community of like-minded professionals who, by the early 1930s,

29 Judith Areschoug, *Det sinnesslöa skolbarnet: Undervisning, tvång och medborgarskap 1925–1954* (Doctoral thesis, Linköping: Linköping's University, 2000), p. 71.

30 While not explored in detail here, it is important to note that the introduction and use of group tests had significant influence on school development in the interwar and postwar periods, both in Norway with Marie Pedersen (above) as the key person, and beyond.

31 Julie Monrad Jacobsen, Jørgen Olsen and Karl Nordlund, 'Nogen indledende bemærkninger om forbundet og tidsskriftet', *Hjælpskolan/Værneskolen/Særskolen*, 1 (1923), 1–4.

32 An example of this arena for exchange was the 1925 meeting in Vendelsberg, Sweden. At this meeting the following members were elected for a three-year period in the board of the Nordic Special School Association: From Denmark, head teacher and president Jørgen Olsen, Sofie Rifbjerg, Asta Sørensen, and Hans Peter T. Ørum. From Norway, chair Julie Monrad Jacobsen, principal John Soyland, head teacher S. Svensen, consultant Otto Grennes, and Anna Stokke. From Sweden, chair Karl Nordlund, Professor Gustav Jaederholm, Anna Tengstrand, head teacher A. V. Hammer, and Ester Gauffin, see Signe Elgenström, 'Nordiska hjälpskolemötet på Vendelsberg folkhögskola den 28–31 juli 1925', *Hjælpskolan/Værneskolen/Særskolen*, 3 (1925), 65–75 (p. 75).

numbered around 500 members—teachers, educators, pedagogues, psychologists, psychiatrists, and medical doctors—united in a common cause: supporting ‘backward’ children and developing special school institutions.<sup>33</sup>

At the ten-day founding meeting, which began on 5 August 1922 in Dalsæter, Norway, with around forty participants, several notable figures and topics emerged, thoroughly reported by Elisabeth Faye-Hansen. Danish psychiatrist Hans Peter Ørum (1877–1950) lectured on the use of IQ tests in Denmark. Norwegian teacher Marta Sørensen (1869–1964) shared her experiences studying IQ testing in the United States. Swedish teacher Anna Tengstrand (1870–1957) spoke about the challenges faced in smaller towns and rural areas, advocating for on-call doctors to administer IQ tests in these peripheral regions.<sup>34</sup> Elisabeth Faye-Hansen (1870–1946) played a central role in the activities of the Nordic Special School Association throughout the 1920s, regularly reporting on its meetings, such as from Dalsæter: ‘After all the talks there was a lively discussion. Most agreed that IQ tests are a good tool for classifying children, but that they cannot stand alone. Cooperation between doctors and educators is crucial’.<sup>35</sup>

Faye-Hansen illustrates the close connections in the “Oslo-circle”. She was a colleague and friend of Anna Sethne (1872–1961). Sethne was head of the Norwegian Female Teacher Association from 1919, editor of its influential journal *Vår Skole*, and principal of the innovative Sagene School in Kristiania/Oslo. Sethne and the Sagene school gained international recognition for progressive, experimental educational practices, attracting visitors from around the world, and particularly from other Scandinavian countries. Sethne’s work garnered both practical and financial support from Oslo school director Bernhof Ribsskog (1883–1963). Together, they were instrumental in establishing

33 By 1933 the association consisted of the following national associations: The Norwegian, Danish and Swedish Special School Associations, which in Sweden had three subdivisions in Stockholm, southern Sweden and Skåne. In 1926, the Special School Association of Finland was founded, and the first Nordic Special School Association meeting was held in Helsinki in 1932, see Julie Monrad Jacobsen, ‘Nordisk Særskoleforbund—Et Tilbakeblikk’, *Hjälpskolan/Värneskolen/Særskolen*, 11 (1933), 73–78 (p. 74).

34 Elisabeth Faye-Hansen, ‘Nordisk Skoleforbunds Møte. En Rekke interessante Spørsmål under Diskusjon’, *Arbeider-Politikken*, 9 August 1922, p. 2.

35 Faye-Hansen, ‘Nordisk Skoleforbunds Møte’, p. 2.

The New School [*Nyskolen*], a model for progressive and exploratory education and reform pedagogy in Norway. Sethne and Ribsskog were leading figures in both the progressive education movement and in Norwegian educational reforms. They contributed significantly to the development, testing, and attempts at standardization of IQ tests in Norwegian schools during the interwar period. Examples include Ribsskog's doctoral thesis *Experimental Contributions to the Psychology of Learning* from 1931 and Sethne's experiments with activity-based learning. Faye-Hansen, likewise, advocated a more flexible school system that could accommodate both gifted and struggling students. In line with contemporary psychometric and psychotechnical thinking, she supported the use of IQ testing to identify slow learners, leading to the designation of 'repeaters' [*gjensittere*], or placement in dedicated special, boarding, or state special schools.<sup>36</sup>

Faye-Hansen, Sethne and Ribsskog were all personally acquainted with Helga Eng, a key figure in the professionalization of pedagogical and psychological disciplines in Norway. She introduced experimental psychology and served as head of what was to become Kristiania/Oslo's psychotechnical institute from the mid-twenties. Eng was Norway's first professor of pedagogy and head of the Faculty of Pedagogy [*Pedagogisk Forskningsinstitutt*] at the *Kongelige Frederiks Universitet* [now University of Oslo] in 1938. Her broad pedagogical interests ranged from children's drawings and art education to child psychology and empirical IQ testing. In 1923, Eng initiated the translation of the Binet–Simon IQ test into Norwegian. At the seventh Nordic meeting on the 'abnormal issue' [*abnormsaken*] in Trondheim in 1926, Eng lectured on 'New Contributions to Child and Youth Psychology'.<sup>37</sup>

Eng maintained a wide professional network, rooted in her extensive travels and collaborations with leading German-speaking psychometricians such as Wilhelm Wundt (1832–1929), William Stern (1871–1938), and Charlotte Bühler (1893–1974).

36 Elisabeth Faye-Hansen, 'Kan skolen gjøres mere elastisk?' *Vår Skole*, 3 (1922), 21–22; Brit Marie Hovland, *Reformpedagogikk på norsk. Anna Sethne og reformskolen på Sagene*, in review.

37 Helga Eng, 'Nyere bidrag til barnets og ungdommens psykologi', *Nordisk Tidsskrift for Dövstumskolan*, 55 (1926), p. 277.



Fig. 2.1 Helga Eng (1876–1966) was the first person to earn a doctorate in psychology in Norway (1913). In 1938, she became the inaugural professor of pedagogy at The Pedagogical Research Institute, Det Kongelige Frederiks Universitet in Oslo. She was a pioneering educationalist, pedagogue, and educational psychologist. Her many books range from test psychology to art pedagogy and they were translated into several languages. Her wide-ranging influence and academic position situate her as an international pioneer within—and between—pedagogy and test psychology in the period the two disciplines diverged. Photographer unknown/ Oslo Museum, CC0 1.0.

In 1925, she visited Gustav Jaederholm (1882–1936), professor of education in Gothenburg and a prominent advocate of IQ testing in Swedish education. Jaederholm, who had studied psychotechnics in the United States and developed IQ tests for Swedish schools, shared his insights with Eng on both American and continental approaches.<sup>38</sup>

In 1925, Eng had been asked to standardize Norwegian tests

38 Elisabeth Lønnå, *Helga Eng: Psykolog och pedagog i barnets århundre* (Oslo: Fagbokforlag, 2002), p. 155.

for Kristiania's vocational schools, when leading what was later institutionalised as the Psychotechnical Institute [*Psykoteknisk institutt*] in 1932.<sup>39</sup> Here they applied "right person in the right place" principles, aiming to match individual abilities and aptitudes with suitable occupations—a psychometric approach that sought to optimize both personal potential and the efficiency of society and the state.

These prosopographical connections highlight the trans-institutional and transnational networks that shaped pedagogical science and the context of introducing IQ testing to the Norwegian school system and academia, not least within the Nordic Special School Association. The association played a central role in bringing together key actors and knowledge brokers from Norway, Sweden, and Denmark. This collaborative spirit was also exemplified by the Hindsgavl meeting in Denmark in 1928, attended by notable proponents of IQ testing, including the Norwegians Marie Pedersen and Johan Lofthus (1887–1959),<sup>40</sup> the Dane Sofie Rifbjerg (1886–1981),<sup>41</sup> and the Swede Signe Elgenström.<sup>42</sup>

## Knowledge Brokers and Knowledge Circulation in Scandinavia

The Nordic Special School Association and its journal served as a platform for exchanging and disseminating knowledge on experiments, practices, courses, and legislative developments across the Nordic countries. In addition, the inter-Scandinavian discourse on IQ testing and concepts of intelligence was enriched through extensive exchanges, including invited talks, lectures, and study trips. In this section, we will examine key examples of these exchanges, analyzing their content to better understand the nature of the interactions and the knowledge

39 See Isak Lønne Emberland, chapter 5, this volume.

40 Lofthus became Norway's first school psychiatrist in 1923 and standardized and adapted the Stanford–Binet test to Norwegian, which was published in 1931. See Lefkadiou, chapter 1, in this volume.

41 Rifbjerg obtained a Master of Arts in psychology in 1925, and she was one of the key actors behind the Danish standardization of the Binet–Simon test, which was published in 1930. In 1936, she joined the educational psychology office in Copenhagen, assuming its leadership in 1938.

42 Elgenström was headmistress of the Stockholm city school for the feeble-minded and editor of the association's journal *Hjälpskolan/Värneskolen/Særskolen*.

related to IQ testing and intelligence that circulated within and into the Scandinavian region.

Gustav Jaederholm, from Sweden, delivered a lecture on 'children's emotional mechanisms' at the Danish Society for Experimental Pedagogy in Copenhagen as early as 1915<sup>43</sup> and was invited to give a talk in the Norwegian Pedagogical Society in November 1925 to present the findings of his book on intelligence measurement in theory and practice.<sup>44</sup>

In preparation for his work standardizing the Stanford–Binet test in Norwegian in 1931, Lofthus visited his Swedish and Danish counterparts in the autumn of 1923. These included psychiatrist Alfild Tamm (1876–1959) from Stockholm, Hans Peter Ørum from Copenhagen, and Professor Christian Keller (1858–1934), head of the mental care institution *De Kellerske Anstalter* in Brejninge, Denmark.<sup>45</sup>

Another conspicuous knowledge broker was the Danish psychologist Henning Meyer (1885–1967), who became the first formally employed education psychologist in Denmark in 1934. Meyer published an article in the Norwegian journal *Skole og Samfund* [School and Society] in 1926, and in November 1934 he appeared, together with Anna Sethne, as a speaker at the university in Oslo for the annual teacher meeting of South-East Norway. He spoke on his work with IQ testing in Frederiksberg and received loud acclamation, according to news reports.<sup>46</sup> In this way, Meyer's work also inspired developments in Norway. At an educational meeting organized by Norderhov and Haug, authorities on IQ testing at school entry, teacher A. Dyrhaug from Bærum cited Meyer's finding that first-grade IQ scores could vary by as much as seven years.<sup>47</sup> From the 1920s onwards, Norwegian educational journals frequently referenced Meyer, and in efforts to establish a faculty of pedagogy at the university, school director Dr. Bernhof Ribsskog pointed to Meyer's achievements in Denmark and similar plans in Sweden.<sup>48</sup>

43 Letter template from the Danish Society for Experimental Pedagogy, 1916, Danish National Archive [Rigsarkivet], Historisk Paedagogisk Studiesamling (HPS), 1.1.1.7 DLH Foreningen for eksperimentel paedagogik I 1913, folder no 1, p. 51.

44 'Dag til dag', *Den Nye Social-Demokraten*, 17 November 1925, p. 4.

45 Ludvigsen, Lundahl and Ydesen, 'Creating an educational testing profession', p. 126.

46 'Østlandsk lærerstevne i Oslo', *Arbeiderbladet*, 31 October 1934, p. 4.

47 'Intelligensprøver ved optagelse på skolen', *Ringerike Blad*, 1 December 1936, p. 1–2 (p. 1).

48 Bergens Arbeiderblad, 'Millioner brukes til skolevesenet, men et forskningsinstitutt

Once again, the neighboring Scandinavian countries served as key references for ongoing educational debates and developments. Numerous other examples of such knowledge brokering in the Scandinavian context can be mentioned. For instance, Ørum was an invited speaker on 'morally defective children' at the Association for Special Schooling in Southern Sweden's course in Gothenburg in January 1924, an association where Jaederholm had been an active and respected board member for several years.<sup>49</sup> Similarly, Riffbjerg gave a talk on "backward children" at the Lillehammer school meeting in 1933.

The exchanges among these knowledge brokers served two primary purposes. Firstly, they learned from one another to stay abreast of the latest developments and practices in neighboring countries, thereby improving their own national practices. One example is Eng's 1925 visit to Jaederholm, undertaken to prepare the establishment of a psychotechnical institute and to test related approaches in Norway. Jaederholm introduced her to the field, offering key contacts and recommendations for her subsequent visits to leading professors on the Continent. Another example of such knowledge exchange is the recurring journal section 'From the Science Lab' in *Hjälpskolan/Värneskolen/Særskolen*, authored by the Danish psychiatrist Ørum.

Secondly, they were committed to advocating for policy improvements within their own countries. For instance, Julie Monrad Jacobsen (1874–1962) passionately argued:

Through their zeal and interest, the normal individuals have asserted their demands, while our poor ones have had only their weakness as their advocate. And this advocate is of little use to society, as it functions more as an unpleasant pang of conscience that one prefers to shake off. However, such unmet demands eventually rise as a judge against society. The wasted lives demand accountability, and society cannot avoid punishment. Shortsighted policies will yield poor socioeconomic outcomes, whereas fair consideration for the demands of the intellectually weak youth would mean labor force gains, savings in public welfare, and even in the prison system. The moral significance is evident.<sup>50</sup>

---

mangler fremdeles', *Bergens Arbeiderblad*, 22 February 1935, p. 1.

49 A.V. Hammer, 'Professor G. A. Jaederholm. In memoriam', *Hjälpskolan/Värneskolen/Særskolen*, (15) 1937, 18–19 (p. 18).

50 Julie Monrad Jacobsen, 'Hvad der kan og bør gjøres for svakt begavet ungdom', *Hjälpskolan/Värneskolen/Særskolen*, 2 (1924), 1–12 (p. 3).



In 1923, Otto Grennes (1868–1943), the director of the Norwegian Schools for the Deaf, Blind, and Intellectually Disabled, similarly lamented:

We have repeatedly and unsuccessfully sought funding from the parliament for a public course for teachers of defective children. Under the current circumstances, there is little hope, if any, of receiving any public funding for such specialized education.<sup>51</sup>

Despite their shared advocacy for intellectually weak children in both educational and societal contexts, Monrad Jacobsen and Grennes articulated this commitment through somewhat different emphases, pointing to distinct strands within contemporary debates on intelligence and IQ testing.

Again, the examples illuminate the emergence of two interconnected knowledge arenas: the pedagogical and educational arena, primarily occupied by educators and special-school teachers, and the scientific, psychological, and medical arena, which was characterized by translators, adaptors, and practitioners of IQ testing. The scientific, psychological, and medical arena emphasized a more strict quantitative measurement approach, guided by the ideal of objective findings and the belief in the hereditary nature of intelligence. In contrast, the pedagogical and educational arena incorporated a larger emphasis on context and professional judgment when utilizing IQ tests.

Ørum and Monrad Jacobsen exemplifies the two “positions”. Ørum represents the scientific, psychological, and medical arena, advocating strongly for strictly scientific arguments based on statistical and quantitative data. While he acknowledged individual differences in phenotypes—describing intelligence as a lumber room [*pulterkammer*]<sup>52</sup>—he remained firm in his belief that the scientific measurement of intelligence was reliable. Ørum asserted that ‘intelligence is entirely dependent on heredity, not the environment, so an increase in general wealth will not raise intelligence,’ a view also shared by Jaederholm. For them, IQ was rooted in biology.<sup>52</sup> On the contrary, Monrad Jacobsen, representing the more pedagogical and educational arena, emphasized

51 Otto Grennes, “‘Den særskilte undervisning’, Foredrag på særskoledagen i Kristiania 29 september 1923’, *Hjälpskolan/Värneskolen/Særskolen*, 1 (1923), 91–102 (p. 93).

52 Hans Peter T. Ørum, ‘Fra Videnskabens Værksted—Arvelighed’, *Hjälpskolan/Värneskolen/Særskolen*, 12 (1934), 44–52 (p. 51); Axelsson, *Rätt elev i rätt klass*, pp. 76, 82.

the following in a presentation on the transfer of children to special schools:

In the deliberations concerning the expulsion of children to special schools, several demands were made: the teacher's experience of the child should be decisive; intelligence measurement, if possible, should be performed by a psychiatrist-trained physician or special class teacher, who has been trained for this purpose by means of special courses; after the pupil's admission to a special school, a thorough physical examination should be carried out.<sup>53</sup>

In contrast to the scientific, psychological, and medical arena, the pedagogical field placed greater emphasis on educational knowledge and judgment over psychometric measurements.

In Norway, the tension between the disciplines of medicine and psychology, and education and pedagogy, was exemplified in a personal communication between school director Grennes and his later successor, Marie Pedersen. In a 1933 letter to her, Grennes expressed deep concern that his position might be inherited by a psychiatrist or medical doctor, thus shifting control away from the pedagogical domain and profession. He wrote:

The management of schools for defective children must lie in the hands of a fully trained educator who also masters the necessary medical subjects relevant to the deaf, blind, and intellectually disabled. [...] I hope that we will be spared from once again seeing a political upstart dealing with these sensitive issues.<sup>54</sup>

The disciplinary and professional contestation was clearly pronounced. As indicated, the knowledge arenas in the interwar inscriptions about IQ testing largely followed professional lines on a continuum between teachers and psychiatrists, with psychologists (and perhaps experimental test-oriented pedagogues) often hovering somewhere in the middle.

The examples reveals the emergence of new academic disciplines and their growing and rivalling influence on both mainstream and special education during a period of reform and reorganization. In Norway,

53 Signe Elgenström, 'Nordiska hjälpskolemötet på Vendelsberg folkhögskola den 28–31 juli 1925', *Hjälpskolan / Värneskolen / Særskolen*, 3 (1925), 65–75 (p. 73).

54 Eva Simonsen, *Vitenskap og profesjonskamp: Opplæring av døde og åndssvake i Norge 1881–1963* (Doctoral thesis, Oslo: Unipub forlag, 2000), p. 345; Dorothee Harms, 'Marie Lovise Pedersen. En Heilpedagogisk pioner i Norge. På sporet av et spesialpedagogisk idégrunnlag' (Master's thesis, Oslo: Universitetet i Oslo, 2015), pp. 4, 43.

Helga Eng, Bernhof Ribsskog, Johan Lofthus, and Marie Pedersen stand out as key persons, each deeply engaged in IQ and maturity testing from their distinct disciplinary and professional positions.

In addition to the association, the journal, and numerous invited talks, a key arena for knowledge exchange and dissemination was the courses held in all three countries to train special education teachers. One notable example is from Denmark, where the country's first special education course began in October 1924.<sup>55</sup> Earlier that year, a course was held by the Association for Special Schooling in Southern Sweden in Gothenburg. Again, we encounter familiar names on the speaker list. Jaederholm planned two lectures; the first was titled 'Psychology and Intelligence Research', and the second was on 'Psychological Views on Didactics and General Education'. Additionally, Monrad Jacobsen addressed the topic 'What Should Be Done for Weakly Gifted Youths?'<sup>56</sup> This course, with about 120 participants from across Sweden, was later reported on by Elgenström, who stated that a Nordic cooperation on joint literature on special schooling was needed.<sup>57</sup> The relationship between professions through academic scholars and special-school teachers was again discussed. She argued that:

The co-operation between the guardians of science and the teachers of the special school is necessary in order for the work to be fruitful on both sides. It was particularly pleasing to note the understanding that existed between the theoreticians of science and the practical workers of the school.<sup>58</sup>

Recognizing the distinct strengths of various professional groups, the journal frequently emphasized the essential collaboration between academic experts and special-school teachers through lectures and meetings. A 1927 article reported that Jaederholm led a course at the University of Gothenburg's psychology department, which included twelve lectures and twenty-four hours of laboratory work, attended by

---

55 Kristjan Haunsø, 'Det første danske Værneskolekursus', *Hjælpskolan/Værneskolen/Særskolen*, 2 (1924), 13–24.

56 Sydsvenska Hjälpsskoleföreningen, 'Provisorisk plan för Sydsvenska Hjälpsskoleföreningens kurs i Göteborg 2–9 januari 1924', *Hjælpskolan/Værneskolen/Særskolen*, 1 (1923), 105–106 (p. 105).

57 Signe Elgenström, 'Från hjälpskolekursen i Göteborg 2–9 januari 1924', *Hjælpskolan/Værneskolen/Særskolen*, (2) 1924, 27–30.

58 Elgenström, 'Från hjälpskolekursen i Göteborg 2–9 januari 1924', p. 29.

nearly all special-school teachers in Gothenburg. These professional groups were united by their shared commitment to advancing special education in the Scandinavian countries.

The network of Nordic educators, psychologists, psychometricians, and psychotechnicians during the interwar period is evident through numerous examples. While different approaches and paradigms regarding intelligence and IQ testing existed, members of the Nordic Special School Association and participants in its activities shared a strong sense of community and a collective mission of great societal importance, particularly for children labeled as backward, their families, and their potential place in society. There was a continuous inter-Scandinavian and transnational exchange of ideas, and a concerted effort to establish best practices.

## Concluding Discussion

This chapter has examined the significance and operational dynamics of central knowledge brokers in a transnational Scandinavian arena focused on special education, with particular emphasis on the development and use of intelligence and IQ testing. It has explored the nature, significance, and implications of interactions and exchanges among key actors and knowledge brokers concerning IQ testing and conceptions of intelligence in Norway, Sweden, and Denmark during the interwar period.

The shared history, linguistic affinity, and mutual perception of these countries as reference societies fostered a high level of exchange, most notably through the establishment of the Nordic Special School Association and its journal *Hjälpskolan/Värneskolen/Særskolen*, whose roots extend back to the nineteenth century.

The Nordic Special School Association and its journal functioned as what Latour has termed a 'centre of calculation'—a hub where actors, practices, and resources from dispersed locations were accumulated, codified, and redistributed. Conferences, training courses, lectures, and study visits acted as circulatory mechanisms, allowing knowledge to be gathered, transformed, and reinserted into national contexts. The collaboration between academic experts and special-school teachers, as highlighted in various journal articles and archival materials, reflects

the existence of a multidisciplinary approach to education.

Our prosopographical approach illuminates two unifying features among Scandinavian knowledge brokers in the interwar period. Firstly, a recognition of common educational challenges in establishing a unitary school system amid rapidly evolving educational sciences and disciplines, which allowed them to draw on practices and knowledge from neighboring countries to improve their own. Secondly, a shared dilemma of designing a reliable, science-based system for determining inclusion and exclusion within education

A key finding of this study is that the development and institutionalization of IQ testing within both normal and special education in Norway, Sweden, and Denmark reflected a dual-system approach. This approach sought to create unitary primary schooling while simultaneously addressing the needs of children deemed unfit for regular education. The emerging unitary education system, later known as the Nordic education model, was founded on a symbiotic relationship between normal and special education. During the interwar years, there was a clear dialectic between inclusion and exclusion in Nordic education, underscoring the need for empirical measurement technologies to objectively identify normality and deviation. IQ tests were seen as essential tools in this effort, with educators, psychologists, psychiatrists, and medical doctors working together to develop an educational system that catered to diverse needs.

However, this comparative Scandinavian analysis has also identified differing paradigms and approaches to intelligence and IQ testing among professional groups, which caused some tension and concern over how to select pupils for special education. Despite these differences, the members of the Nordic Special School Association shared a strong sense of mission, fostering a community united by the goal of influencing national policies. Their ambition was to address both societal and individual needs, ensuring that all children received appropriate and effective schooling for their own benefit and for the benefit of society.

The tensions and aspirations traced here remain relevant in contemporary debates over inclusive education. The problem of balancing universal provision with targeted differentiation, while avoiding stigma, continues to animate educational policy discussions both within the Nordic countries and globally. Historical reflections on

the Scandinavian experience show how the negotiation of such tensions has long been embedded in transnational exchanges, professional boundary work, and the politics of measurement.

This issue remains relevant today, both globally and within the context of the Nordic education model, highlighting the ongoing challenge of balancing inclusion and differentiation. For instance, the debate over how to delineate between normal, special, and mental care education, while avoiding stigmatization, mirrors contemporary discussions on inclusive education.

In conclusion, the transnational Scandinavian arena played a pivotal role in shaping the development of IQ testing practices and concepts of intelligence in Norway, Sweden, and Denmark, influencing educational policies and curricular reforms. This historical analysis provides valuable insights into the formative years of IQ testing and its lasting impact on educational systems in Scandinavia.

## Bibliography

- Areschoug, Judith, *Det sinnesslöa skolbarnet: Undervisning, tvång och medborgarskap 1925–1954* (Doctoral thesis, Linköping: Linköping's University, 2000).
- Axelsson, Thom, *Rätt elev i rätt klass: Skola, begåvning och styrning 1910–1950* (Doctoral thesis, Linköping: Linköping's University, 2007), <https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-7994>
- Beech, Jason and Alejandro Artopoulos, 'Interpreting the circulation of educational discourse across space: Searching for new vocabularies', *Globalisation, Societies and Education*, 14 (2016), 251–271, <https://doi.org/10.1080/14767724.2015.1025713>
- Claparède, Edouard, 'Helga Eng, Experimentelle Untersuchungen über das Gefülsleben des Kindes im Vergleich mit dem des Erwachsenen', *Archives de Psychologie*, 18 (1923).
- 'Dag til dag', *Den Nye Social-Demokraten*, 17 November 1925, p. 4.
- Elgenström, Signe, 'Från hjälpskolekursen i Göteborg 2–9 januari 1924', *Hjälpskolan/Värneskolen/Særskolen*, 2 (1924), 27–30.
- Elgenström, Signe, 'Nordiska hjälpskolemötet på Vendelsberg folkhögskola den 28–31 juli 1925', *Hjälpskolan/Värneskolen/Særskolen*, 3 (1925), 65–75.
- Eng, Helga, 'Nyere bidrag til barnets og ungdommens psykologi', *Nordisk Tidsskrift for Døvstumskolen*, 55 (1926), 277.

- Faye-Hansen, Elisabeth, 'Kan skolen gjøres mere elastisk?' *Vår Skole*, 3 (1922), 21–22.
- Faye-Hansen, Elisabeth, 'Nordisk Skoleforbunds Møte. En Rekke interessante Spørsmål under Diskusjon', *Arbeider-Politikken*, 9 August 1922, p. 2.
- Grennes, Otto, "'Den særskilte undervisning'", Foredrag på særskoledagen i Kristiania 29 september 1923', *Hjälpskolan/Värneskolen/Særskolen*, 1 (1923), 91–102.
- Hammer, A.V., 'Professor G. A. Jaederholm. In memoriam', *Hjälpskolan/Värneskolen/Særskolen*, 15 (1937), 18–19.
- Hamre, Bjørn, 'Den yderste halespids: Bekymring som dispositiv i skolens problemforståelser og forskelssætninger', *Uddannelseshistorie*, 46 (2012), 70–94.
- Hamre, Bjørn, Thom Axelsson and Kari Ludvigsen, 'Psychiatry in the Sorting of Schoolchildren in Scandinavia 1920–1950: IQ Testing, Child Guidance Clinics, and Hospitalisation', *Paedagogica Historica*, 55 (2019), 391–415, <https://doi.org/10.1080/00309230.2019.1597137>
- Harms, Dorothee, 'Marie Lovise Pedersen. En Heilpedagogisk pioner i Norge. På sporet av et spesialpedagogisk idégrunnlag' (Master's thesis, Oslo: Universitetet i Oslo, 2015).
- Haunsø, Kristjan, 'Det første danske Værneskolekursus', *Hjälpskolan/Värneskolen/Særskolen*, 2 (1924), 13–24.
- Haunsø, Kristjan, 'Hvad skal vi kalde skolen for de svagt begavede?', *Hjälpskolan/Värneskolen/Særskolen*, 12 (1935), 8–14.
- Hovland, Brit Marie, *For fedreland og broderband: Allmugeskulelærarane si møteverksemd i det nasjonale gjennombrøtet 1850–1870* (Volda: Høgskolen i Volda, 1999).
- Hovland, Brit Marie, 'Mapping the Nation—Fabricating Citizens from Patriotism Towards Nordicism?' in *The Nordic Education Model in Context*, ed. by Daniel Tröhler, Bernadette Hörmann, Sverre Tveit and Inga Bostad (New York: Routledge, 2022), pp. 15–35, <https://doi.org/10.4324/9781003218180-3>
- Hovland, Brit Marie, *Reformpedagogikk på norsk. Anna Sethne og reformskolen på Sagene*, in review.
- 'Intelligensprøver ved optagelse på skolen', *Ringerike Blad*, 1 December 1936, p. 1–2.
- Jónasson, Jón Torfi and Valgerður S. Bjarnadóttir, 'The Nordic interactive and comparative spaces within the arena of education', *Nordic Studies in Education*, 43 (2023), 25–42, <https://doi.org/10.23865/nse.v43.3987>

- Jöns, Heike, 'Centre of calculation', in *The SAGE Handbook of Geographical Knowledge*, ed. by John Agnew and David N. Livingstone (London: SAGE Publications Ltd, 2011), pp. 158–170, <https://doi.org/10.4135/9781446201091.n13>
- Kaalund-Jørgensen, Frederik C. 'Hvad gør vi for de Børn, der ikke kan følge Folkeskolens almindelige Undervisning?', *Hjælpskolan/Værneskolen/Særskolen*, 20 (1942), 87–95.
- Kettunen, Pauli, 'The transnational construction of national challenges: The ambiguous Nordic model of welfare and competitiveness', in *Beyond Welfare State Models: Transnational Historical Perspectives on Social Policy*, ed. by Pauli Kettunen and Klaus Petersen (Cheltenham: Edward Elgar, 2011), pp. 16–40, <https://doi.org/10.4337/9781849809603.00008>
- Krejsler, John Benedicto, 'The Nordic dimension as a metaspace for educational research', *Nordic Studies in Education*, 43 (2023), 8–24, <https://doi.org/10.23865/nse.v43.5421>
- Lawn, Martin, 'Voyages of Measurement in Education in the Twentieth Century: Experts, Tools and Centres', *European Educational Research Journal*, 12 (2013), 108–119, <https://doi.org/10.2304/eerj.2013.12.1.108>
- Ludvigsen, Kari, Christian Lundahl and Christian Ydesen, 'Creating an educational testing profession—the emergence and impact of the Scandinavian testing community, 1920–1960', *European Educational Research Journal*, 12 (2013), 120–138, <https://doi.org/10.2304/eerj.2013.12.1.120>
- Lundahl, Christian, 'Making testers out of teachers: The work of a Swedish State Research Institute 1946–1956', *History of Education*, 48 (2019), 646–63, <https://doi.org/10.1080/0046760X.2019.1565422>
- Lønnå, Elisabeth, *Helga Eng: Psykolog och pedagog i barnets århundre* (Oslo: Fagbokforlag, 2002).
- 'Millioner brukes til skolevesenet, men et forskningsinstitutt mangler fremdeles', *Bergens Arbeiderblad*, 22 February 1935, p. 1.
- Moljord, Gøril and Hilde Bondevik, 'Constructions of the In-/Educable', in *The Nordic Education Model in Context*, ed. by Daniel Tröhler, Bernadette Hörmann, Sverre Tveit and Inga Bostad (New York: Routledge, 2022), pp. 56–74, <https://doi.org/10.4324/9781003218180-5>
- Monrad Jacobsen, Julie, 'Hvad der kan og bør gøres for svakt begavet ungdom', *Hjælpskolan/Værneskolen/Særskolen*, 2 (1924), 1–12.
- Monrad Jacobsen, Julie, 'Nordisk Særskoleforbund—Et Tilbakeblikk', *Hjælpskolan/Værneskolen/Særskolen*, 11 (1933), 73–78.
- Monrad Jacobsen, Julie, Jørgen Olsen and Karl Nordlund, 'Nogen indledende bemærkninger om forbundet og tidsskriftet', *Hjælpskolan/Værneskolen/Særskolen*, 1 (1923), 1–4.



- Mülberger, Annette, 'The need for contextual approaches to the history of mental testing', *History of Psychology*, 17 (2014), 177–186, <https://doi.org/10.1037/a0037487>
- 'Nordisk Værneskolemøde', *Decorah-Posten*, 21 August 1928, p. 3.
- People's School Act [Lov om folkeskolen i Kjøstæderne/paa Landet], 1889 (Norway) (Stockholm Slot/Kristiania: P.T. Mallings Boghandels Forlag, 1890).
- Popkewitz, Thomas S., 'Comparative reasoning, fabrication, and International Education Assessments: Desires about nations, society, and populations', *International Journal of Educational Research*, 112 (2022), <https://doi.org/10.1016/j.ijer.2022.101940>
- Simonsen, Eva, *Vitenskap og profesjonskamp: Opplæring av døve og åndssvake i Norge 1881–1963* (Doctoral thesis, Oslo: Unipub forlag, 2000).
- Simonsen, Eva, "'Folkeskolens Ytterste Forpost"—En Annen Historie Om Den Norske Folkeskolen 1945–1970', *Uddannelseshistorie*, 26 (2012), 118–44.
- Skard, Åse Gruda, 'Arbeidet ved en mentalhygienisk skuleklinikk', *Norsk pedagogisk årbok* 7 (1934–35), 75–80.
- Skard, Åse Gruda, *Psykologi og psykologar i Norge* (Oslo: Oslo Universitetsforlaget, 1959).
- Statens offentliga utredningar, 1936:31, *Betänkande med utredning och förslag angående åtgärder för särskild undervisning och utbildning av psykiskt efterblivna barn i barn- och ungdomsåren* (Stockholm: Ivar Haeggströms boktryckeri A.B, 1936).
- Sydsvenska Hjälpsskoleföreningen, 'Provisorisk plan för Sydsvenska Hjälpsskoleföreningens kurs i Göteborg 2–9 januari 1924', *Hjälpsskolan/Værneskolen/Særskolen*, 1 (1923), 105–106.
- Thuen, Harald, *Den norske skolen: Utdanningssystemets historie* (Oslo: Abstrakt forlag, 2017).
- Thuen, Harald, *Om barnet: Oppdragelse, opplæring og omsorg gjennom historien*. (Oslo: Abstrakt forlag, 2008).
- Tröhler, Daniel, 'Introduction: The Nordic education model', in *The Nordic Education Model in Context*, ed. by Daniel Tröhler, Bernadette Hörmann, Sverre Tveit and Inga Bostad (New York: Routledge, 2022), pp. 1–11, <https://doi.org/10.4324/9781003218180-1>
- Verboven Koenraad, Myriam Carlier and Jan Dumolyn, *A Short Manual to the Art of Prosopography*. In *Prosopography Approaches and Applications. A Handbook* (Oxford: Unit for Prosopographical Research, Linacre College, 2007), pp. 35–69, <https://hdl.handle.net/1854/LU-376535>

- Weber, Matthew S. and Itzhak Yanovitzky, 'Knowledge Brokers, Networks, and the Policymaking Process', in *Networks, Knowledge Brokers and the Public Policymaking Process*, ed. by Matthew S. Weber and Itzhak Yanovitzky (New York: Palgrave, 2021), pp. 1–25, [https://doi.org/10.1007/978-3-030-78755-4\\_1](https://doi.org/10.1007/978-3-030-78755-4_1)
- Ydesen, Christian, *The Rise of High-Stakes Educational Testing in Denmark, 1920–1970* (Doctoral thesis, Frankfurt a.M: Peter Lang Verlag, 2011).
- Ydesen, Christian, 'The Shadowy Existence of the Nordic Dimension in Education', *Nordic Studies in Education*, 43 (2023), 260–265, <https://doi.org/10.23865/nse.v43.5674>
- Ørum, Hans Peter T., 'Fra Videnskabens Værksted—Arvelighed', *Hjælpskolan/Værneskolen/Særskolen*, 12 (1934), 44–52.
- 'Østlandsk lærerstevne i Oslo', *Arbeiderbladet*, 31 October 1934, p. 4.



### 3. A Legal History of Intelligence, Testing, and Criminal Unaccountability

*Linda Gröning and Svein Atle Skålevåg*

---

In criminal law, in Norway as in many other countries, “intellectual disability” is understood to reduce or negate certain capacities needed for criminal accountability. According to Section 20 of the Norwegian Penal Code, a high degree of intellectual disability can make defendants unaccountable for otherwise criminalized acts.<sup>1</sup> Forensic experts (psychiatrists or psychologists) evaluate defendants and inform judges about the mental state and disabilities of the defendant. In this work, they rely on, among other things, intelligence testing.<sup>2</sup> This practice has its roots in the early twentieth century.

When intelligence testing was first introduced, criminal law was suggested as a field where these tests could prove useful. Alfred Binet and other early proponents of intelligence testing in the beginning of the twentieth century pointed to criminal law as a field where the tests could potentially do good service to society.<sup>3</sup> Indeed, psychiatrists working in the field as legal experts were quick to pick them up. The first Norwegian adopters of intelligence tests also pointed to criminal law

---

1 The Norwegian Penal code (LOV-2005-05-20-28). Available in English at <https://lovdata.no/pro/#document/NLE/lov/2005-05-20-28>. For an explanation of the rule in section 20, see Linda Gröning, ‘Criminal Insanity in Norwegian Law’, in *The Insanity Defense. International and Comparative Perspectives*, ed. by Ronnie Mackay and Warren Brookbanks (Oxford University Press, 2022), pp. 295–315. See also below in the section about the law reform of 2020.

2 See further about the role of the experts in Norwegian law in Gröning 2022, p. 312–314. See further Victoria Westrum and Linda Gröning, ‘Strafferettslig utilregnelighet ved psykisk utviklingshemming—en diskusjon om regelen og praktiseringen av den.’, *Tidsskrift for strafferett*, 3 (2022), 214–237.

3 Alfred Binet, ‘Preface’, *L’année psychologique*, 14.1 (1908).

as a field of application of this new tool.<sup>4</sup> They offered promises that intelligence testing could make it possible to turn the normative question of accountability, at least partially, into an empirical question that could be answered objectively, with the help of professional testers. In a word borrowed from recent studies in the history of science, the tests seem to have promised ‘mechanical objectivity’ in criminal law.<sup>5</sup> This noble dream of objectivity has since lingered on in the legal life of intelligence testing.

At the same time, intelligence testing in criminal law has hardly been a topic of historical analysis. Historical studies have primarily been concerned with intelligence testing of children in an educational context.<sup>6</sup> Consequently, little is known about how intelligence tests became relevant in criminal law and how their relevance has changed over time. In this chapter, we will explore and discuss how the introduction of intelligence testing, and the concept of intelligence that was associated with it, came to affect legal reasoning about the unaccountable criminal in Norwegian legal history. The chapter is based on legal and forensic documents. The key question is how the introduction of intelligence tests and the concept(s) of intelligence that came with them have influenced the legal understanding and construction of the (un)accountable offender. By addressing this question, we aim to contribute, on the one hand, to the literature on criminal accountability, where the history of intelligence tests has not been much studied; and on the other hand, the literature of intelligence testing, where the use of these tests in the setting of criminal law has not been in focus.

The chapter is structured chronologically, taking major legal reforms of statutory rules about criminal (un)accountability as reference points. We will start from the discourse that surrounded the enactment of the accountability rule in the Norwegian penal code of 1842 and follow the developments up to the current penal code.

4 S. Dahlstrøm, *Forsøk med Binet–Simons intelligensprøver* (1922); C. Looft ‘Intelligensundersøkelser av skolebarn’, *Medicinsk Revue*, 30.10 (1913), 569–585.

5 For histories of objectivity, see Daston and Galison, *Objectivity* (Zone Books, 2007); T. M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton University Press, 1995); Peter Novick, *That Noble Dream. The “Objectivity Question” and the American Historical Profession* (Cambridge University Press, 1988).

6 See e.g., J. Carson, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750–1940*, (2007); A. Mülberger, ‘Biographies of a Scientific Subject: The intelligence test’, in *Oxford Research Encyclopedias: Psychology* (Oxford University Press, 2022); A. Mülberger, M. Balltandre and A. Graus, ‘Aims of Teachers’ Psychometry: Intelligence Testing in Barcelona (1920)’, *History of Psychology*, 17.3 (2014), 206–222, <https://doi.org/10.1037/a0033339>

## Sources and Perspectives

To answer our key questions about intelligence tests and concept(s) of intelligence in the legal understanding of the (un)accountable offender, we apply a qualitative text analysis. Our source material consists mainly of preparatory legislative documents underlying the legal rules on criminal accountability and forensic psychiatric reports.<sup>7</sup> Our historical analysis has been centered around significant legal reforms that have occurred from the 1840s to today. Through exploration of legal and forensic reasoning at the different times of these reforms, our focus is on identifying general changes over a longer period of time.

In our analysis of these documents, we combine approaches from the disciplines of law and history, addressing criminal law as a conceptual landscape. We are interested in the meanings that are immanent in the way the words are used in our source material. An important premise is that the sources we study have been produced by different agents, within different legal and forensic institutions, to serve different functions. Legal rules are enacted to be followed by the citizens and used by legal agents; preparatory legislative documents are written to provide reasons for rules. Forensic reports are addressed to the legal authorities, especially judges, to illuminate criminal proceedings. The forensic psychiatric reports are not, strictly speaking, part of the legislative process, but nevertheless an important part of the foundation of legal discourse.<sup>8</sup>

A challenge in this kind of analysis that stays close to the words used, is how to translate the Norwegian words into modern English without losing the historical resonance of these words. As described in the introduction of this book, it is a general challenge to understand historical material with current language as our departure point. Key constructs in this chapter are, as explained above, “intellectual disability” and

---

7 For this paper we have collected a sample of ca. 200 forensic psychiatric reports from the archive of The Board of Forensic Medicine (*Den rettsmedisinske kommisjon*), which refers explicitly to the use of intelligence testing, from the years 1913–1940. We can assume that this sample constitutes the majority of the reports from this period in which intelligence testing was used. A separate paper with findings from this study will be forthcoming.

8 For a further historical analysis of the significant role of forensic psychiatry in Norwegian criminal law, see S. A. Skålevåg, *Utilregnelighet: En historie om rett og medisin* (Pax forlag, 2016).

“criminal unaccountability”. However, while criminal unaccountability works as a reference term for our historical analysis, intellectual disability is more difficult to use in such a way. Throughout history, many other labels have been used, such as “idiots” or “feeble-minded”, as descriptions relating to, but not necessarily equivalent to, the current label “intellectual disability”. Many of these terms were originally clinical terms that later would become insults in popular language.

To avoid losing important nuances, we will here utilize the historical terms throughout, and seek to explain their historical meaning. We will abstain from commenting on whether the historical terms today appear derogatory or offensive, but leave such reflections to the reader. In our translation of key terms throughout the chapter, we have attempted to stay true to the original terms, and we have also referenced the original language in the notes. When official translations have been available, we have utilized these. In addition, to the extent possible, we have tried to utilize common language, rather than specific disciplinary terms, which has also sometimes been necessary to bridge our different historical and legal perspectives.

Our analysis of the application of intelligence testing in criminal law includes contextual exploration of the concept of intelligence itself. A valuable idea in the historiography of intelligence testing is, in this regard, what we may call ‘the reification thesis’, that is, the view that the historical importance of intelligence testing lies in it contributing to making intelligence a “thing”, and that an important reason why this happened was that it served the professional self-interests of those aspiring to cultural recognition as a profession. Simon Jarrett writes about the IQ as the ‘scientific colonization of mental deficiency’ in his recent book about the disabled mind: ‘As it [the IQ] was taken up enthusiastically in Europe and the United States, intelligence somehow became ‘a thing’ that the power of psychology and its practitioners could extract from people’s minds in perfectly measured form.’<sup>9</sup> Though we consider the reification thesis an important dimension of intelligence testing, our emphasis on the contextual function of the tests will nonetheless bring some needed nuances to this thesis. Our exploration

---

9 S. Jarrett, *Those they Called Idiots: The Idea of the Disabled Mind from 1700 to the Present Day* (Reaktion Books, 2020). A similar idea was proposed by Steven J. Gould, see S. J. Gould, *The Mismeasure of Man* (Norton, 1981).

will show that the introduction and use of IQ tests in criminal law cannot easily be attributed to the desire of forensic experts or lawyers to exercise power.

## Reason and Freedom in the 1840s

Was there a concept of intelligence in legal discourse before the introduction of intelligence testing? To understand this, we need to start before proper intelligence tests emerged, and direct our attention to the criminal code of 1842. There is no direct mention of intelligence in the discourse around the rule of accountability in this code. Still, as we will see, the conceptualization of the unaccountable person may have paved the way for the intelligence tests as legally relevant at a later stage.

The Norwegian penal code of 1842 was produced in the tradition of the French *code penal* of 1810, and the Bavarian penal code of 1813.<sup>10</sup> This code was a major event in the life of the young Norwegian nation, as it replaced the heterogenous collections of laws and ad-hoc prescripts from the age of the absolute monarchy. It also introduced a general rule of criminal accountability: 'Acts that are committed by the furious and the demented, or by those that are deprived of the use of reason from disease or the weakness of old age shall go unpunished'.<sup>11</sup>

Implied in this rule is a specific conception of a person, the basis of which are constituted by reason and freedom. The underlying assumption was that an accountable person is someone who has reason, and the freedom to follow this reason.<sup>12</sup> The conditions mentioned—furor, dementia, weakness due to old age or disease—all have in common the fact that they were conditions considered to exclude reason. The rationale for this rule is found in the legal motives and legal discourse more broadly. Reason was, from the legal-philosophical point of view, an absolute category, something one either had or did not.<sup>13</sup> It was what

10 R. Hauge, *Straffens begrunnelse* (Universitetsforlaget, 1996).

11 'De handlinger er straffrie, som forøves av galne eller afsindige eller af dem, som Forstandens brug ved Sygdom eller Alderdoms-svaghet er forøvet.' Ch 7, §2

12 The code of 1842 belonged in a legal discourse concerned with acts, not agents, hence the code spoke of acts that an offender should not account for, and should go unpunished—and not directly about unaccountable persons.

13 Carson, *The Measure of Merit*, p. 4; L. Holberg, *Introduction til Natur- og Folkeretten* (2013).



distinguished humans from animals, but also, in some exceptional cases, an attribute that one might lose, becoming unreasonable and thus hardly human.

At the same time, a recognition of ‘gradual accountability’ emerged in the context of judicial sentencing. The legal motives also acknowledged that the acts of defendants may be *more* or *less* accountable, and hence the severity of the sentence should be proportionate to the degree of reduced accountability. Among aspects that could reduce accountability, the motives especially mentioned the defendant’s upbringing, maturity, and also, notably, their “*sjelsevner*”.<sup>14</sup> The meaning of the latter is not evident, but the word would come to play an important role in the Norwegian history of intelligence and criminal accountability. We therefore need to pause to understand what this term might have meant in 1842.

The term *sjelsevner*, which literally might translate to ‘capabilities of the soul’, may have been a novelty in a legal context, but it was not a complete neologism in ordinary language.<sup>15</sup> It occurred occasionally in the Dano-Norwegian language, though not frequently, in the first decades of the nineteenth century. When it occurred, it was predominantly in a pedagogical or a medical context.<sup>16</sup> Authors discussing the mental development of children might use it, as might physicians discussing mental illness.<sup>17</sup> One author in 1818 discussed whether it was possible to stimulate the development of the *sjelsevner* in education. Another author in 1835 enlisted the various capabilities that were relevant to this construct—‘reason, judgment, cleverness, wit and recollection’.<sup>18</sup> The examples indicate that the predominant meaning in these contexts is the “mind”, and a more exact translation of the expression would

14 Lovcommissionen av 1828, *Motiver til det i aaret 1832 udgivne Forslag til en Lov for kongeriget Norge angaaende Forbrydelser* (Christiania, 1835). The relevant passage is from chapter six, on fixing of sentences.

15 Note that the written language in Norway in the nineteenth century was Danish. Only at the very end of the century did Danish and Norwegian develop into two distinct languages.

16 Since the term is relatively unusual, we have performed a search in the database of the Norwegian National Library (nb.no), which contains most Norwegian newspapers, periodicals, and published books. We found thirty-three occurrences in newspapers in the period 1801–1842, and twenty-seven occurrences in periodicals.

17 Rektor Rosted, ‘Innbydelsesskrift til den offentlige eksamen ved Christiania Kathedralskole’ (1818); Anonymous, ‘Responsa Medica’in *Eyr*, no. 2 (1827).

18 ‘Forstand, Dømmekraft, Skarpsind, Vid og hukommelse kort sagt alle hendes Sjeleevner’; Anonymous, ‘Misceller’, in *Bien: Et maanedsskrift for moerskabslæsning*, vol 12, nr 3 (1835), 122–124.

be 'mental capabilities' (and we will therefore use this translation henceforth). As for the legal discourse of the 1840s, what is important is that this vague concept seemed to open up an understanding of the mind as an entity measured in degrees, as opposed to the dichotomy of reason and unreason.

For practicing judges, the thinking expressed in the motives of the code of 1842 meant that they would have to form an opinion about the mental state of the defendant before them, and that to a certain extent they were expected to adjust the punishment to the nature of the criminal as well as to the nature of the crime. None of the sources relating to the rule in the 1842 code say anything about seeking expert advice, but internationally, the alienist or psychiatrist was already emerging as the courts' preferred expert on human nature. Even though the reference to mental capabilities was not taken from medical discourse, it does seem to have created a room into which a medical expert could and did move. As such, the remark about mental capabilities in legislative motives is an important moment in this history, as it signals the need for a psychiatric assessment in order for the law to function properly at a time when it was rather uncommon in Norway for any medical professional to actually act as an expert in the courts.<sup>19</sup>

The reference to the mental capabilities of the defendant as something that could be measured in degrees seems to have allowed for a connection between criminal law and psychiatric assessments of these capabilities. This connection was reinforced through the discourse around the criminal code from 1902, to which we now turn.

## Impaired Capabilities in the Early 1900s

Where the 1842 code was colored by Enlightenment sensibilities, the criminal code of 1902 was conceived under the influence of the new international school of positivist criminology (especially F. von Liszt, G. van Hamel and A. Prins).<sup>20</sup> On certain points it was also influenced by a medical profession, and an emerging corps of psychiatrists, that had become significantly more powerful in the decades since the first criminal code was adopted. Some of these psychiatrists shared many concerns

19 This had happened, but it was still very rare. See Skålevåg, *Utilregnelighet*.

20 R. Hauge, *Straffens begrunnelse* (Universitetsforlaget, 1996).

with positivist criminology. Basic elements of positivistic thinking, broadly speaking, were that criminal law should integrate scientific insights about human nature and avoid metaphysical assumptions. Moreover, the positivists advocated for a view of criminal law as an instrument in the governance of society, as opposed to a discourse for expressing blame on behalf of society.<sup>21</sup> To the adherents of positivist criminology, and similar contemporary networks, the criminal was to be seen as a distinct human type, with a characteristic psychology. Among the characteristics of this criminal psychology could be an arrested or deviant development.<sup>22</sup>

What is relevant for us is that the code of 1902 included a revised statutory rule about criminal unaccountability that gave *sjelsevner* a new meaning. To this end, the code also introduced reactions that were not, strictly speaking, punishments, but rather measures directed towards prevention of serious crimes committed by offenders considered to be dangerous.

The rule about criminal unaccountability followed from section 44 and had the following wording:

An act is not punishable if the person acting at the time of its commission was insane, unconscious or otherwise unaccountable due to insufficient development of the mental capabilities or impairment or pathological disturbance of these or due to coercion or imminent danger.<sup>23</sup>

Now, there was no longer any explicit reference to reason in the rule. Instead, the rule included an explicit reference to *sjelsevner*. As we have explained above, the expression 'mental capabilities' had a history both inside legal discourse and outside of it, but it seems to have taken on a

21 P. Becker, P. and R.F. Wetzell, eds., *Criminals and Their Scientists: The History of Criminology in International Perspective* (German Historical Institute, Cambridge University Press, 2006); Michel Foucault, 'About the concept of the "Dangerous individual" in 19th century legal psychiatry', *International Journal of Law and Psychiatry*, 1 (1978), 1–18; Hauge, *Straffens begrunnelse*

22 This use of "human kinds" is from Ian Hacking, 'Degeneracy, Criminal Behavior, and Looping', in *Genetics and Criminal Behavior*, ed. by D. Wasserman and R. Wachbroit (Cambridge University Press, 2001), pp. 141–168, <https://doi.org/10.1017/CBO9781139173162.006>

23 'En handling er ikke straffbar, naar den handlende ved dens foretagelse var sindssyg, bevidstløs eller iøvrigt utilregnelig paa grund af mangelfuld udvikling af sjælsevnerne eller svækkelse eller sygelig forstyrrelse af disse eller formedelst tvang eller overhængende fare'.

new meaning in 1902. The intellectual historian Espen Schaanning has argued that this word in 1902 became associated with a new concept of the human, one that emphasized psychological drives and passions rather than the faculty of reason. According to the new concept, the mind could be imagined as insufficiently developed, impaired, or pathologically disturbed. This transformation has been seen as indebted to developments in the discourse of mental medicine.<sup>24</sup> As such, the 1902 rule, and the new meaning of mental capabilities, signaled a significant shift in criminal law towards a recognition of the offender as a complex psychological being with capabilities and impairments that could be studied and measured empirically.

Yet, this does not mean that the reference to reason disappeared completely in the discourse surrounding the code of 1902. The earlier understanding of mental capabilities, as relating to human reason, continued to haunt it.<sup>25</sup> This unstable concept of mental capabilities, mixing the reasonable and the instinctual, was important when the intelligence tests were introduced in criminal law, a decade after the introduction of the code of 1902. A closer examination of the early use of intelligence testing is needed to understand what the tests did to the conceptual landscape of the law. In the following section, we will provide a closer look at the forensic psychiatric reports from the 1910s and 1920s. We will return to the reports from the 1930s in a later section.

## Intelligence Testing in Forensic Practice in the 1920s

The first documented instance of formal intelligence testing as part of the forensic psychiatric examination in Norway, is from the case of an arsonist in 1913. It is noteworthy that this was the same year that also saw the first Norwegian medical publication about intelligence testing.<sup>26</sup>

24 Espen Schaanning, *Kampen om den forbryterske sjel. Kriminal-filosofiske vitenstrekk*. (Unipubforlag, 2002).

25 See e.g. NOU 1974: '17 Strafferettslig utilregnelighet og strafferettslige særreaksjoner', p. 217, which quotes Langfeldt 1959, claiming that in the first years after 1930, many legal professionals had difficulties in understanding that 'diminished mental capacities' did not mean diminished intelligence.

26 DRK 61/1914. (The number refers to DRK (the Board of forensic medicine), the journal number and the year). For a description of the sample, see footnote 8, above. The first Norwegian medical publication that discussed intelligence testing was a book review by Sigurd Dahlstrøm, reviewing Charles Dana's book *Mental Testing*, in

A forensic report was based on a meeting between the physician/psychiatrist in the role of an “observer”, and the charged person in the role of an “observee”. These labels reflected the empirical ethos of the forensic practice, and more specifically the idea of the psychiatrist as a passive observer. In this particular case, the observee was found, by means of the Binet–Simon tests, to have the mental age of an eight-year-old. But the tests were only a small part of the general examination (or observation). From the broad examination, the experts concluded that the observee was “*sinnssyk*” [insane] and represented a danger to the public.

The examination was conducted by the doctors Ragnar Vogt and Paul Winge. Vogt was the first, and at this time only, professor in psychiatry in the country. Winge, on the other hand, was the first to hold a PhD in forensic psychiatry. Both had a seat in *Den rettsmedisinske kommisjon* [the Norwegian Board of Forensic Medicine]. The Board was a government-appointed committee of medical experts, established in 1900, to guarantee that the work of forensic experts around the country was of a satisfactory scientific standard.<sup>27</sup> A copy of every report in forensic medicine presented to courts of law or prosecutors around the country was submitted to the board. The members would then signal any inadequacies in the reports. An important reason for establishing this board and investing in it a function of such control was the fact that many of the physicians that acted as legal experts in the Norwegian province had very little training in forensic psychiatry.<sup>28</sup> The board had two divisions, one for psychiatry and one for pathology. Winge and Vogt together made up the psychiatric division, and from this position they dominated forensic psychiatry in the country until their deaths in 1920 and 1943 respectively.

---

*Norsk magasin for lægevidenskaben*, Vol. 5, no. 11, pp. 1420–1435.

27 S. A. Skålevåg, ‘A culture of consensus: Organising expertise in Norwegian forensic psychiatry, late nineteenth to early twentieth century’, in *Forensic Cultures in Modern Europe*, ed. by Willemijn Ruberg, Lara Bergers, Sara Serrano Martínez, Pauline Dirven (Manchester University Press 2022).

28 In practice, the physicians acting as experts in the legal system were not necessarily certified psychiatrists. We do nonetheless refer to all the professionals who acted as psychiatric experts in the legal system as (de facto) psychiatric experts.

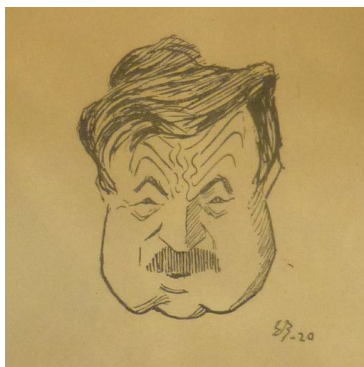


Fig. 3.1 Caricature of Professor Ragnar Vogt (1870–1943). Vogt studied in Oslo, Leipzig, Heidelberg and Copenhagen. He earned his doctoral degree in 1901 on research in experimental psychology, and in 1915 Vogt became the first professor of psychiatry in Norway. He sat for many years as a member of the Board of forensic psychiatry. Photo: Norsk Teknisk Museum, CC BY-NC-SA.



Fig. 3.2 Paul Winge (1857–1920) studied medicine in Oslo, and earned his doctoral degree in 1896 on a dissertation titled ‘The forensic medical examination of the insane criminal’. He was employed as a police and prison physician from 1893 to his death in 1920. In this role he encountered and examined a large number of criminals, probably more than any other Norwegian physician of his time. He was instrumental in establishing a Norwegian Board of Forensic Medicine, and sat as one of its psychiatric members for twenty years. Photographer L. Forbech 1901/National Library of Norway, PDM 1.0.

In addition to their strategic position in the board, Winge and Vogt also frequently served as first line experts for the courts, as in the above-mentioned case of the “insane” arsonist. Their involvement in this case shows how intelligence testing was introduced by the very elite within the emerging forensic psychiatric system.

Intelligence tests were used, but only infrequently, by forensic psychiatrists in the following years. In the late 1910s, they were only used by members of a small group of experts in Oslo, and predominantly in dealing with arsonists suspected of being “idiots”, as in the case introduced above.<sup>29</sup> From this time, the well-established term “idiot”, referring to those with profound intellectual disabilities, came to be understood in light of the intelligence scale. The standard procedure of intelligence testing was that the appointed forensic psychiatrist himself conducted the tests. In only one case in the material were the tests performed by someone who held a junior position, as a psychiatrist in training under the forensic expert, functioning as an aide to the expert.<sup>30</sup> In none of the cases that we have studied, however, were the tests performed by a non-physician.

From 1920, a growing number of the persons tested were charged for sexual offenses, such as *blodskam* [incest], even though arsonists continue to be very present in the material.<sup>31</sup> From this time, experts from outside the Oslo group also started to test defendants that they were appointed to examine with formal intelligence tests.

In the psychiatric reports from the period of their introduction and through the 1920s, the tests are regularly talked about in the plural and referred to as a ‘method’.<sup>32</sup> The observees do not ‘take’ the test, but are examined ‘with’ the tests, or ‘after’ the tests (i.e. they are examined after the procedure prescribed by the tests). In at least one case, the observee is said to be ‘submitted’ to the test.<sup>33</sup> In another case, the experts report to have used Binet–Simon’s *undersøkelsesskjema* [form]”.<sup>34</sup> The result of

29 DRK 61/1914, 62/1914, 152/1917, 204/1919, 112/1920. The experts were P. Winge, R. Vogt, J. Scharffenberg and S. Dahlstrøm.

30 DRK 62/1920, also DRK 241/20.

31 Sexual offences: DRK 261/1920, 51/1921, 249/1924, 249/1924, 277/1925, 232/1925. Arsonists after 1920: DRK 9/1928, 11/1928, 29/1928, 94/1928, 133/1928, 143/1928, 158/1929.

32 DRK 261/1920.

33 DRK 29/28.

34 DRK 222/28.

the examination of a defendant's intellectual functioning was generally reported as an age "level" [*trinn*], as seen above, when an observee is functioning at 'the level of an eight-year-old'.<sup>35</sup> In some cases, this is reported as 'intelligence age'. A report from 1916 refers to a "so-called" intelligence age of 9.3, indicating that the concept of intelligence age was still not very well known.<sup>36</sup> Other examples refer not to an overall score, but to the various tasks included in the tests, as in a case from 1917 where the experts found that the observee managed well 'the tasks which a normally developed child at the nine to ten-year level can do'.<sup>37</sup>

The reference to a 'level' may be seen as a prolongation of differentiation according to the school system. Before there was an intelligence test, the schools would sort people according to their age level, and in certain cases a pupil would have to remain in a class when their peers moved on. In Binet and Simon's manual, the different tests were also organized according to level. Such an implicit reference to the levels of the school system seems to have been abandoned with the introduction of the intelligence quotient (IQ), a number that places the observee on a continuous scale, instead of a more crude level. In the forensic psychiatric reports, the concept of IQ is first found in a report from 1928. This report was written by Johan Lofthus, who later earned his PhD in medicine for his thesis on intelligence testing.<sup>38</sup> Lofthus reports the collected result of the intelligence test as both IA (11 10/12 year) and IQ (74). In other cases, the report would not (only) give a number as the result of the collected tests, but detail the results of the different tests, listing how many of the tests on each level the observee got right.<sup>39</sup>

In some cases, the rhetoric of the report seems to indicate that the intelligence tests functioned as a "mechanical" tool, i.e. a tool that produces an exact answer with no need for interpretation of the results. This was the case when the experts reported an intelligence age of 9 years, equating an IQ of 0.56, and therefore found the observee to belong to 'the class of feeble-minded individuals that are designated debile'.<sup>40</sup> But more often, the expert seems to be careful not to signal

35 E.g. DRK 62/1914.

36 DRK 184/1916. See also DRK 51/1921.

37 DRK 152/1917.

38 DRK 143/1928. The other expert in this case was Kristensen, a police physician.

39 DRK 112/1920, 261/23.

40 DRK 222/1928. 'Vi antar derfor at observanden tilhører den klasse åndssvake



such a mechanical objectivity in the report. This is the case in a report from 1925 where the expert finds that the observee belongs at the seven-year level, or 'possibly slightly higher', measured after Binet-Simon's "method". As the expert concludes that this is a case of 'idiotia', he emphasizes that the result of the tests is confirmed by the impression he has otherwise formed of the observee.<sup>41</sup>

We should mention here that the classifications of intellectual disability that the forensic psychiatrists evoked in the 1920s (idiots, imbeciles, debiles), did not have a place in the legal code, nor were they discussed in legal documents.<sup>42</sup> As we have shared, the legal code recognized the insane and people with deficient mental capabilities. As far as the forensic psychiatrists were concerned, the idiots formed a distinct class of people within the larger category of the insane. As for the less disabled, those referred to as 'debiles', the legal implications were far less clear.

All in all, the corpus of forensic reports from the 1910s and '20s documents how Norwegian forensic psychiatrists started to integrate the Binet-Simon tests in their examinations of defendants in criminal cases. It shows no sign that the tests were pushed by professionals outside the community of forensic psychiatrists, and they do not show signs of an enthusiastic embrace of a new method that could give their work the allure of objectivity. In the 1910s and 1920s, the tests were not used very frequently in the practice of Norwegian forensic psychiatrists. An explanation may be that the forensic experts were not yet familiar with the procedure. It may also be the case that there was a certain reluctance among the experts towards employing a tool that was developed for children on adults. Paul Winge, for instance, remarked in a report in 1916 that he found the tests generally less valuable for adults than for children.<sup>43</sup> The psychiatrist Harald Arnesen repeated the same criticism in 1918, in a published text strongly critical of the Norwegian forensic

---

individer som betegnes som debile.' The term feeble-minded is here used in its broadest sense, encompassing both idiots, imbecils and debiles.

41 'Dette sammenholdt med det inntrykk jeg ellers har fått av hans åndsevner mener jeg berettiger meg til å stille diagnosen idiotia', DRK 232/1925. See also DRK 349/1932, where the expert is reluctant to put too much emphasis on the tests, as he has not found other indicators of mental illness (other than an intelligence age of 8-9).

42 See e.g. Straffelovskommissionen, *Udkast til almindelig borgerlig Straffelov for Kongeriget Norge II* (Motiver, 1896).

43 DRK 154/16

psychiatric establishment.<sup>44</sup> At the same time as there were such critical voices, the intelligence tests, and the conceptions about intelligence these tests produced, had started to find their way into the legal discourse.

## Intelligence Testing in the Legal Discourse in the Late 1920s

In 1929, the criminal code of 1902 was revised in a way that affected and consolidated the role of intelligence testing in Norwegian criminal law. First, the statutory rule on unaccountability was rephrased, so that it came to align more directly with the medical system. Second, the now familiar term of mental capabilities was moved within the criminal code, away from the rule about criminal responsibility, instead associated with rules on dangerous criminals.<sup>45</sup>

The revision of the code was the result of a political process that started in 1922, and took place in the background of a public debate on sexual offences.<sup>46</sup> In brief, there was significant public pressure to legislate for more severe punishments for these crimes. Several women's organizations fought a veritable campaign to make this happen, beneath the banner of the protection of children and women. Among the new punishments they wished for was the castration of sexual offenders. The intensified attention on sexual crimes was accompanied by more intense problematization of the relationship between human nature and legal transgression. It would result in a sterilization law, but also in a revision of the criminal code.<sup>47</sup>

In 1922, a law commission was appointed by the government to prepare the ground for reform of the criminal code, most importantly of the rules on criminal responsibility.<sup>48</sup> As was (and is) the custom, the commission consisted of experts and stakeholders whose task was to produce a consensus report with recommendations to the

---

44 H. Arnesen, 'Sindsyk—aandsvak. Den dominerende retning i norsk psykaitri', *Tidsskrift for den norske Lægeforening*, 38(24) (1918), 1049–1059.

45 Skålevåg, *Utilregnelighet*.

46 S. A. Skålevåg, 'Kjønnsforbrytelser. Sedelighet, seksualitet og strafferett 1880 - 1930', *Tidsskrift for kjønnsforskning* (1–2) (2009), 7–27.

47 A sterilization law, partly based on the principles of eugenics, was enacted in 1934. See Haave and Kyllingstad, chapter 4, this volume.

48 Skålevåg, *Utilregnelighet*.

government. Among the experts were no less than three physicians, among them Ragnar Vogt, professor of psychiatry. There were also two female physicians, Ingeborg Aas and Tove Mohr, who may have been appointed as stakeholders representing the women's movement. All these physician members represented something of a breakthrough for a medical perspective on criminal law. The commission delivered their report in 1925.<sup>49</sup> After further legislative processes, the recommendations in the report eventually led to the enactment of a revised rule on criminal accountability, effective from 1929.

The rule about criminal accountability was revised to the following wording: 'An act is not punishable if the perpetrator was insane or unconscious when it was committed'.<sup>50</sup>

The revision introduced a clearly articulated 'medical model' for criminal unaccountability. It now exclusively defined unaccountability by the conditions of being 'insane' or 'unconscious', and without requiring any assessment of how the defendants' condition influenced the commission of the crime.<sup>51</sup> The new wording of the rule was also more simplified in respect to previous rules.

As such, the 1929 revision shows the influence of positivism at that time, in its avoidance of questions regarded as metaphysical, such as questions of guilt and free will. The new code also carefully avoided asking for an assessment of how mental states influenced the crime. Within the framework of criminological positivism, such an assessment of a causal link between a mental state and an act was considered a metaphysical question.<sup>52</sup> It was a rule designed with the aim of being practicable first of all, and it facilitated a clear division of labor between psychiatric and legal professionals. It should come as no surprise that the formulation had first been proposed by the Norwegian Medical

49 Innstilling fra den av Justisdepartementet 11. mai 1922 opnevnte komité til revisjon av Straffeloven (1925).

50 'En handling er ikke straffbar, når gjerningsmannen ved dens foretagelse var sinnssyk eller bevisstløs'.

51 S. A. Skålevåg, 'Kjønnsforbrytelser. Sedelighet, seksualitet og strafferett 1880 - 1930', *Tidsskrift for kjønnsforskning* (1-2) (2009), 7-27. For a discussion about the medical model tradition in criminal law, see also L. Gröning, 'Has Norway abandoned its medical model? Thoughts about the criminal insanity reform post 22 July', *Criminal Law Review*, 3 (2021), 191-202.

52 For a critical discussion of such perspectives in criminal law, see Gröning, 'Tilregnelighet og utilregnelighet: Begreper og regler', attachment to the official criminal law commission report NOU 2014: 10, 422-426.

Association in 1900, as they pushed for a satisfactory organization of the work of forensic experts.<sup>53</sup>

The notion of *sinnsykdom* [insanity] became the preferred term of criminal law in 1929. In our context, however, it is significant that this criterion was also understood to include those who, according to the terminology of the time, were labelled “idiots”, i.e. those who later, as we will see, were labelled as intellectually disabled. It was therefore relevant for the 1922 commission to define an “idiot”, and they did so by referring to the new metric scale for intelligence, and the concept of intelligence age that was associated with it. The threshold for unaccountability was set by the commission to the intelligence age of eight years old.

Thus, it was implied that intelligence testing was relevant for criminal law, but the commission did not specifically discuss the tests, their validity or their reliability. The IQ that had been introduced by Stern and later Terman in the 1910s was never explicitly mentioned. In fact, the word “intelligence”, in this important report, was never invoked independently from the measuring of it. This seems to support the view that intelligence as a concept was introduced in legal reasoning on the back of the intelligence tests.

Another important change through the revisions of 1929 was that the term “mental capabilities” was removed from the accountability rule. We have seen that this amorphous expression entered criminal law in 1902, as a factor to reckon with when assessing the accountability of an act. In 1929, it was taken out of the rule about unaccountability, and instead included in a new rule about preventive sanctions for the most dangerous criminals. If defendants were found to have impaired mental capabilities and there was a real danger that they might repeat their offenses, then they were liable to *særreaksjoner* [special reactions], which could be, for example, preventive detention or a ban on alcohol consumption. It was common for these preventive sanctions to not be seen as punishments for an offender, strictly speaking, but as measures for the protection of public safety, in continuation of the points of

---

53 *Betænkning afgivet af en af Den norske lægeforening nedsat komité, angaaende de bestemmelser i straffelovudkastet, der særlig interesserer lægestanden. Stortingsforhandlinger 1900–1901, 5 d, Dokument nr 1.*

view of the positivistic school.<sup>54</sup> Such a broadening of the spectrum of protective measures was also an important background for the sterilization law of 1934.

When it came to identifying people with impaired mental capabilities, it was generally assumed that the assistance of forensic psychiatric experts was necessary. Surprisingly, the commission did not discuss the meaning of “mental capabilities” to any degree, implicitly leaving it to the community of forensic psychiatric experts to define it.

The paradigmatic dangerous offender, in this new regime, was the sexual offender, who was an offender that more often than not was understood to act in accordance with his nature. The preferred expert on human nature was the medical expert. In this way, we may say, the revisions in 1929 admitted an expanded scope for the competency of the psychiatric expert: not only were they asked to assist in the identification of the defendant’s accountability for the criminal act, but also to identify the offender who represented a specific threat to society by their very nature.

This larger role of the expert is visible in an appendix to the 1925 report, which encompasses a series of seventeen case histories written by the committee member Ragnar Vogt. These case histories served to illuminate the paradigmatic status of the sexual offender, and the relationship between what was considered mental diseases and certain crimes (specifically sexual crimes).<sup>55</sup> Seven of these seventeen case histories dealt with “idiots”, and seven of them dealt with persons with *svekkede sjelsevner* [impaired mental capacities] who were considered to be mentally impaired to a lesser degree than the “idiots”. In at least eleven of the cases, the forensic psychiatrist used an intelligence test to assess the intelligence age of the defendant, though none of the case histories give any details about the testing practice. Thus, the case histories provide evidence to the reader that intelligence age was a relevant factor in criminal law, and that intelligence testing was part of the forensic psychiatric practice. In practice, there is little doubt that the forensic psychiatric experts before 1929 were already trusted to assist

54 See e.g. A. Prins, *La Défense sociale et les transformations du droit penal* (1910).

55 Innstilling fra den av Justisdepartementet 11. mai 1922 opnevnte komité til revisjon av Straffeloven (1925): 167–182

in the identification of dangerous offenders. Nonetheless, the reform of 1929 consolidated and formally anchored this function in statutory legal rules, and this played a part in the consequent increase in the use of forensic psychiatric experts in criminal cases about unaccountability and dangerousness.

For a great many of the psychiatric experts that frequently informed criminal proceedings, the concept of intelligence was significant for their understanding of intellectually disabled offenders. Here, an assessment of the offender's mental age would prove to be a persistent element in the expert's assessment of the defendant's capabilities.

### Intelligence Testing in Forensic Practice in the 1930s

Beginning in 1930, we see a significant rise in the frequency of intelligence testing in Norwegian forensic psychiatric reports. This seems to have been a direct consequence of the legal changes in 1929, effective from 1930. From this time, the forensic psychiatrist was customarily asked two questions: first, whether the defendant was insane (psychotic or intellectually disabled). Second, whether the defendant's *sjelsevner* were impaired or insufficiently developed, in case preventive sanctions could be needed. In some cases, they would also be asked whether there was a danger of future reoffending. In forming an opinion, the experts increasingly made use of the intelligence tests of Binet, Simon, and Terman. It is not clear from the legal and forensic sources that we have studied which versions of the tests were used.

We see in the available reports from the 1930s that intelligence tests were used more frequently than in the previous decade, that they were used by a larger number of experts, and that they were used in the examination of a wider spectrum of defendants.

As already pointed out, in 1930, "mental capabilities" still had a rather unstable status and meaning. This instability is visible in the fact that some forensic psychiatrists avoided the expression altogether, preferring linguistic variations such as *aandsevner* [spiritual abilities], instead of *sjelsevner* [mental capabilities]. There existed no available "test" or "method" to identify these capabilities, only an emerging professional consensus on how to assess it, and in this assessment, intelligence remained a significant part.

In the 1930s, as in the 1920s, only in very few cases did the experts embrace the idea of an IQ, even if they were clearly aware of the existence of this concept. In most cases, they held on to the more original concept of an intelligence age, and even with regards to this concept, they often expressed an approximative approach.<sup>56</sup> In many cases, it is more accurate to say that the observees were observed by an expert *using* the Binet–Simon tests, than to say that they *took* a test.<sup>57</sup> Used in this way, it could be argued, psychiatrists amplified the importance of their professional judgment—a hallmark of the medical profession. They refused to be reduced to operators of mechanical objectivity, and instead used the tests to bolster their judgment.

## IQ Scores in Legal Discourse from the 1970s

Almost fifty years passed before the rules on criminal accountability in the 1970s were subjected to new reform proposals. During the 1970s, 1980s, and 1990s, there were several such proposals, but it was not until 2002 that the criminal code was actually changed. As the proposals leading to this change tell us something interesting about conceptions of intelligence and the legal relevance of IQ tests, we will here describe some central aspects.

The reform work in the 1970s started with the appointment of *Straffelovrådet* (the permanent criminal law commission) in the early 1970s. Among the commission's ten members were four psychiatrists, proving that the medical perspective was still seen as highly relevant for criminal law. *Straffelovrådet* delivered their report in 1974, and this set off the process towards a new penal code that continued in the 1980s and 1990s.<sup>58</sup>

The commission suggested a rule that established 'mentally disordered' and 'intellectually disabled' as distinct criteria for unaccountability: 'A person who, at the time of the act, was seriously mentally disordered (insane) or mentally retarded to a significant

56 See for a further overview and discussion about the understanding of unaccountability and IQ in legal practice, Westrum and Grøning 2022.

57 This is in line with Carson's observation: 'The subject would not so much *take* the Binet–Simon scale as be *observed* by an expert using it to fashion a diagnosis'. Carson, *The Measure of Merit*, p. 141.

58 NOU 1974: 17 Strafferettslig utilregnelighet og strafferettslige særreaksjoner.

degree (feeble-minded) is not punished. A fine can still be imposed'.<sup>59</sup>

As the commission remarked, the legal criteria for the mentally disordered had, in Norwegian practice, previously always included the "mentally retarded" offenders. Yet the commission found it useful, at this point, to separate these as two different conditions regulated in different criteria. As for the intellectually disabled, the commission proposed a more fine-grained terminology, based on available psychiatric literature. The experts in the commission now opted for 'oligophrenia' as the overarching term, and subdivided it into the "debile", the "imbecile", and the "idiot".<sup>60</sup> What is important in this context is that the category oligophrenia was considered a gradual condition, and that the tool to measure it was the IQ tests: 'The degree of oligophrenia is primarily measured with intelligence tests, the result of which is expressed in an IQ value'.<sup>61</sup> The debile had a score of 75–56, the imbecile 36–55, and the idiot below 36.

In addition, the commission suggested adding a facultative clause, allowing the courts (but not obliging them) to see another group of offenders as also unaccountable, from the borderland of mental disorder and retardation. The commission referred to this group as people with severe conditions of abnormalities. In this group, we also find the group of "debiles", that is people with an IQ of 56–75.<sup>62</sup> These people could, under certain further specified circumstances, also be considered unaccountable.

It is noteworthy that the commission made reference to the intelligence scale as a seemingly objective standard. At the same time, the commission acknowledged that the threshold for the different legal categories was based on mere psychiatric convention, which had shifted over time. The commission specifically referred to the practice of forensic

59 NOU 1974: 17, s. 149. 'Den som på handlingstiden var alvorlig sinnslidende (sinnssyk) eller psykisk utviklingshemmet i betydelig grad (åndssvak) straffes ikke. Bot kan likevel ilegges' (§ 44).

60 The term oligophrenia is here used in the same broad meaning as was previously given feeble-minded, see note 41. Previously the term "feeble-minded" had at times referred to the whole class of intellectually disabled (below IQ 75), and at times only to the most severely disabled (IQ below 55).

61 'Graden av oligofreni males først og fremst ved intelligens tester hvis resultat kommer til uttrykk i en IQ-verdi (Intelligenskvotient)'. NOU 1974: 17, p. 45.

62 This observation is backed by a reference to a textbook in forensic psychiatry, which shows the degree of integration between legal and psychiatric discourses. NOU 1974: 17 Strafferettslig utilregnelighet og strafferettslige særreaksjoner, p. 61.



psychiatry, where those with an IQ score below fifty had previously been considered as suffering from a significant degree of intellectual disability and were criminally unaccountable. For a period of time, the same threshold had been set to IQ 35, while a score between 35 and 55 would call for an individual assessment based on criteria other than IQ tests. In more 'recent times', notes the commission, the threshold had 'again' been set to 50.<sup>63</sup> All this serves as a reminder that even if the IQ scale was considered objective, the demarcation line between the different constructs of the 'idiot' and the 'imbecile', or between the 'imbecile' and the 'debile' was normative.

None of the changes proposed were effectuated. Instead, a new commission was appointed in 1985, called *Særreaksjonsutvalget*, that delivered its report, "NOU 1990: 5", in 1990.<sup>64</sup> This commission proposed both a revised unaccountability rule and a new system for preventive reactions for dangerous offenders. The 1990 commission here reiterated the suggestion from 1974 to have a rule with separate legal criteria for the psychotic and the intellectually disabled offenders, demarcated by a guiding IQ limit of 55.

The 1990 commission is notable for being the first in this series to extensively discuss the international diagnostic manuals—the Diagnostic and Statistical Manual of Mental Disorders (DSM) and the International Classification of Diseases (ICD). Norway now committed in international conventions to build on these manuals in psychiatric practice, which in turn influenced the legal understanding of criminal unaccountability. This reliance on diagnostic manuals, and the relevance of IQ limits in them, came to strengthen the trust in "objective" forensic tools. At the same time, the commission noted that one should not rely too much on formalized IQ tests, but that other personality traits and social functioning should also be included in the overall legal assessment of the defendants' (un)accountability.<sup>65</sup>

The proposal from the 1990 commission resulted in the enactment of a new unaccountability rule in 1997. This rule was finally enacted

63 NOU 1974: 17 Strafferettslig utilregnelighet og strafferettslige særreaksjoner (p. 45, see also p. 3).

64 NOU 1990:5 Strafferettslige utilregnelighetsregler og særreaksjoner: Straffelovskommisjonens delutredning IV.

65 NOU 1990: 5, p. 13.

in 2002, with the following wording:<sup>66</sup> ‘Anyone who was psychotic or unconscious at the time of the act is not punished. The same applies to those who, at the time of the action, were mentally retarded to a high degree’.<sup>67</sup>

This rule was continued when the new criminal code of 2005 was enacted.<sup>68</sup> It survived for almost ten years, until a new major law reform was initiated in the aftermath of the case concerning Anders Behring Breivik, who killed seventy-seven people in Oslo and Utøya on 22 July 2011, and whose trial triggered a major public debate on the role of forensic psychiatrists in Norwegian criminal law.<sup>69</sup>

## The Legal Discourse in the 2020s: An Ambivalence Towards the IQ

The most recent law reform was initiated with the appointment of a new law commission in January 2013.<sup>70</sup> This commission collected a broad range of competence, with representatives from law, medicine, and psychology, and even specialists in philosophy and ethics. With the Breivik case as its backdrop, the commission was asked to review the rule on criminal unaccountability with special emphasis on the criterion of “psychotic”, and to perform a comprehensive assessment of the role of forensic experts.

The commission presented its recommendations in the report NOU 2014:10.<sup>71</sup> Contrary to political expectations, the commission proposed to preserve the medical model and the criterion ‘psychotic’, but also to include conditions that are ‘equally serious’ as being psychotic.<sup>72</sup>

---

66 Endringslov til straffeloven mv. av 1997.

67 ‘Den som på handlingstiden var psykotisk eller bevisstløs straffes ikke. Det samme gjelder den som på handlingstiden var psykisk utviklingshemmet i høy grad’.

68 Ot.prp.nr.90 (2003–2004) s. 423.

69 See for a discussion and critique, Thomas Frøberg, ‘De nye reglene om strafferettslig skyldene’, *Lov og Rett*, 3 (2020), 167–184.

70 See for an explanation of the mandate and conclusions of this commission, Grønning and Rieber-Mohn, ‘Proposal for New Rules Regarding Criminal Insanity and Related Issues, Norway post-22 July’, *BJCLCJ* (2015), 109–131, <https://boap.uib.no/index.php/BJCLCJ/article/view/830>. One of the authors of this article was a member of this committee.

71 Excerpts of the report are available in English at <https://www.regjeringen.no/en/dokumenter/nou-2014-10/id2008986/>.

72 NOU 2014:10, pp. 123–131, 138–140.

The commission proposed no changes in the criterion concerning intellectual disability. But it proposed an addition to this criterion, including those who were *tilsvarende svekket* [correspondingly weakened] as those with severe intellectual disability.<sup>73</sup> The addition was meant to include those who were not intellectually disabled according to the medical ICD-10 diagnosis because they became disabled as adults, for instance as a result of a brain injury.<sup>74</sup> This shows that the commission now connected the legal criterion of intellectual disability directly to the ICD-10 diagnosis that required the disabilities to be apparent in childhood.<sup>75</sup> Thus, similar to the 1990 commission, this commission referred to the ICD system in their explanations of legally relevant intellectual disability, and here confirmed the relevance of a diagnostic IQ scale.<sup>76</sup> Like the 1990 commission, the 2014 commission stated that the legal criteria of intellectual disability were not entirely overlapping with the ICD-10 diagnoses. The commission proposed to continue the guiding limit at IQ 55 or lower, which, according to the diagnostic categories in ICD-10, includes mild to profound intellectual disabilities. Even stronger than the 1990 commission, the commission emphasized that measuring the IQ was not sufficient for assessing criminal unaccountability, but that personality and social functioning also mattered to this legal demarcation. The commission referred to a recommendation from the Norwegian Board of Forensic Medicine to supplement formal IQ tests (Wechsler Adult Intelligence Scale, WAIS) with neuropsychological tests in borderline cases for unaccountability. Here, the commission explained that judges must be aware that the results of different tests can vary.<sup>77</sup>

It is also significant that the commission explicitly sought to reduce the influence of the forensic experts on legal judgements. According to the commission, the experts should no longer comment on whether the defendant's condition satisfies legal criteria, as they were entrusted to before, according to the mandates from courts.<sup>78</sup> Instead, it said

73 The category 'severe intellectual disability' was proposed to be demarcated by a guiding IQ limit at 55, although an overall assessment of functioning in the end should be decisive.

74 NOU 2014: 10, p. 130.

75 NOU 2014: 10, p. 105–106.

76 NOU 2014: 10, p. 105–106.

77 NOU 2014: 10, p. 52.

78 See NOU 2014:10, pp. 216–218.

that experts should merely inform the judge about the diagnoses and functional impairments of the defendant, leaving it to the judge to assess whether the legal criteria for unaccountability were satisfied. As we will see, such limitations do not hinder the continuation of a practice with IQ testing of intellectually disabled defendants, nor do these tests influence legal judgments about criminal (un)accountability.

The law proposal from the government followed in 2017.<sup>79</sup> The government built on the report from the commission regarding the role of the experts, but proposed an unaccountability rule that was markedly different from the commission's proposal. Contrary to the century-old medical model tradition in Norwegian criminal law, the government proposed to avoid the explicit use of medical terminology in the rule on unaccountability, in favor of a rule that provided larger room for juridical discretion.<sup>80</sup> According to the government, the new rule reflected a 'modified mixed model', although it should still not be the mandate of the experts nor the court of law to assess whether the defendant's condition influenced the commission of the crime.<sup>81</sup> The new rule came into force in 2020, in section 20 of the criminal code of 2005, and has the following wording:<sup>82</sup>

A person that at the time of committing the act is under 15 years old is not criminally liable. The same applies to a person who at the time of the act is not accountable due to a

- a) Severely deviant state of mind,
- b) Severely impaired consciousness or
- c) Severe mental disability

When deciding whether a person is not accountable according to the first section, a-c, emphasis shall be to the persons' degree of failure in perception of reality and functional capacity.

79 For a commentary on this proposal, see Grønning, 'Hvordan skal vi avgjøre om alvorlig sinnslidelse innebærer utilregnelighet? Refleksjoner om lovforslaget i Prop. 153 L', *Bergen Journal of Criminal Law and Criminal Justice*, 5 (1), (2016–2017), 77–85, <https://doi.org/10.15845/bjclj.v5i1.1384>.

80 Prop. 154 L (2016–2017), pp. 12–14.

81 Prop. 154 L (2016–2017), pp. 7, 13, 55. See for a discussion about this change, Grønning, 'Has Norway abandoned its medical model? Thoughts about the criminal insanity reform post 22 July', *Criminal Law Review*, 3 (2021), 191–202.

82 This is the official English translation provided by lovdata.no.

Criminal unaccountability now requires both that the defendant's condition fulfils one or more of the criteria specified in the second paragraph letters a–c, *and* that the defendant is 'unaccountable due to' this condition, understood as a matter of 'failure in perception of reality and functional capacity'. Thus, this rule clearly differs from the previous rule that only had criteria referring to specified medical conditions.

The criterion for intellectually disabled offenders in (c) relies, as before, on diagnostic categories of intellectual disability, and it includes different types and levels of intellectual disability.<sup>83</sup> However, the threshold for this criterion is now set by a guiding IQ limit at 60, and not as before at 55.<sup>84</sup> This change was explained by the government, with reference to a psychiatric point of view, in that today's society may require a higher IQ to navigate than was the case before.<sup>84</sup> At the same time, according to the new rule, the court must also independently assess whether a defendant with severe intellectual disability is 'unaccountable due to this condition'. It is explained in the government's proposal that this requirement is meant to downplay the significance of IQ in favor of a stronger emphasis on functional abilities.<sup>85</sup>

Hence, the new rule, and the proposal underlying it, communicates a certain ambivalence to the legal status of IQ tests. On the one hand, it continues and consolidates the reliance on IQ and IQ tests in the demarcation of unaccountability for intellectually disabled offenders. On the other hand, it emphasizes the need for a more comprehensive assessment of the defendant's functioning than the IQ tests alone can provide, and a larger scope for judicial discretion.

This ambivalence is also still seen in legal practice. Since the enactment of the new rule, there have been variations in forensic and legal assessments of intellectually disabled offenders. There are even some cases where the guiding IQ limit still seems to be interpreted as an absolute demarcation, so that formalized IQ tests can determine who can be considered unaccountable and who cannot. In fact, it seems like the application of the criminal unaccountability rule still allows for IQ to be approached as an "objective" and clear threshold for the judges in their difficult judgments of responsibility and punishment.

---

83 Prop. 154 L (2016–2017), p. 229.

84 Prop. 154 L (2016–2017), pp. 89, 229.

85 Prop. 154 L (2016–2017), p. 13.

## Conclusion: Intelligence Tests as Pragmatic Tools

Our historical exploration shows that intelligence tests, and the concepts of intelligence that came with them since the 1910s, have influenced the legal understanding of the unaccountable offender. Intelligence testing and corresponding concepts of intelligence entered the legal discourse when the emphasis of reason in the law was diminished in favor of a positivistic notion of measurable mental capabilities. Interestingly enough, it seems that the concept of intelligence has not been of any legal interest before these tests entered the legal scene. Hence, it is reasonable to say that the intelligence tests brought a concept of intelligence into the law and legal discourse. Over time, the deliberations over the question of unaccountability came to lean increasingly on the intelligence scale—but only belatedly, from the 1970s, was this scale explicitly expressed as an intelligence *quotient* (IQ) in the forensic and legal discourses. In the 1990s, the legal relevance of IQ was strengthened by the integration of the ICD system in the legal reasoning about the meaning of the unaccountability rule. From this time, a guiding IQ limit was utilized to define the threshold for criminal unaccountability.

Thus, it seems like intelligence tests have contributed to a metamorphosis of the legal definitions of unaccountability of intellectually disabled offenders, from normative ideas of lack of reason to specifications of measurable impairments in mental capabilities. In short, the IQ has indeed contributed to a reification of intelligence in criminal law too. But our study also suggests that the tests never quite delivered on the promises of an objective identification of unaccountable offenders.

What seems clear from our exploration of the legal and forensic sources is that the intelligence tests, within the context of criminal law, have played a role in the sometimes difficult interaction between law and psychiatry. At this point, our analysis may add some nuances to the reification thesis introduced in the beginning of the chapter. On the one hand, IQ tests functioned as tools for transforming the mind of the unaccountable offender into a measurable thing, as expressed by the reification thesis. On the other hand, our analysis indicates that legal as well as forensic experts expressed ambivalence and reservation towards the tests and the numbers they produced. Both forensic psychiatrists

and lawyers were hesitant and tentative in their use of intelligence tests. Yet the tests functioned to legitimize the very role of the forensic expert, bolstering their claim to objectivity and reliability. As such, these tests also found their way into criminal law as tools to help the legal demarcation of accountability, which in turn seemingly provided authority to lawyers in addressing this challenging issue.

Ultimately, a reification of intelligence was more an effect of the interaction between law and psychiatry than a strategic instrument in a power struggle. It was driven more by concrete practical needs than by power interests.

## Bibliography

- Arnesen, Harald, 'Sindsyk - aandsvak. Den dominerende retning i norsk psykiatri', *Tidsskrift for den norske Lægeforening*, 38.24 (1918), 1049–1059.
- Becker, Peter, and Richard Wetzell, eds., *Criminals and Their Scientists: The History of Criminology in International Perspective* (German Historical Institute, Cambridge University Press, 2006).
- Binet, Alfred, 'Preface', *L'année psychologique*, 14.1 (1908).
- Binet, Alfred and Théodore Simon, 'Le développement de l'intelligence chez les enfants', *L'année psychologique*, 14 (1908), 1–95.
- Carson, John, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750–1940* (Princeton University Press, 2007), <https://doi.org/10.2307/j.ctv346qgf>
- Dahlstrøm, Sigurd, *Forsøk med Binet–Simons intelligensprøver* (1922).
- Daston, Lorraine, and Peter Galison, *Objectivity* (Zone Books, 2007).
- Foucault, Michel, 'About the concept of the "Dangerous individual" in 19th century legal psychiatry', *International Journal of Law and Psychiatry* 1 (1978) 1–18, [https://doi.org/10.1016/0160-2527\(78\)90020-1](https://doi.org/10.1016/0160-2527(78)90020-1)
- Foucault, Michel, *Les Anormaux. Cours au Collège de France. 1974–1975* (Gallimard Le Seuil, 1999).
- Frøberg, Thomas, 'De nye reglene om strafferettslig skyldene', *Lov og Rett*, 59.3 (2020), 167–184, <https://doi.org/10.18261/issn.1504-3061-2020-03-05>
- Gould, Steven Jay, *The Mismeasure of Man* (Norton, 1981).
- Grønning, Linda, Erling Johannes Husabø and Jørn Jacobsen, *Frihet, forbrytelse og straff* (Fagbokforlaget, 2023).

- Grønning, Linda, 'Criminal Insanity in Norwegian Law' in *The Insanity Defense. International and Comparative Perspectives*, ed. by Ronnie Mackay and Warren Brookbanks (Oxford University Press, 2022), pp. 295–315, <https://doi.org/10.1093/oso/9780198854944.003.0013>
- Grønning, Linda, 'Hvordan skal vi avgjøre om alvorlig sinnslidelse innebærer utilregnelighet? Refleksjoner om lovforslaget i Prop. 153 L', *Bergen Journal of Criminal Law & Criminal Justice*, 5.1 (2016–17), 77–85, <https://doi.org/10.15845/bjclcj.v5i1.1384>
- Grønning, Linda and Georg Fredrik Rieber-Mohn, 'Proposal for New Rules Regarding Criminal Insanity and Related Issues, Norway post-22 July', *Bergen Journal of Criminal Law & Criminal Justice*, 3.1 (2015), 109–131, <https://boap.uib.no/index.php/BJCLCJ/article/view/830>
- Hacking, Ian, 'Degeneracy, Criminal Behavior, and Looping', in *Genetics and Criminal Behavior*, ed. by D. Wasserman and R. Wachbroit (Cambridge University Press, 2001), pp. 141–168, <https://doi.org/10.1017/CBO9781139173162.006>
- Hauge, Ragnar, *Straffens begrunnelse* (Universitetsforlaget, 1996).
- Holberg, Ludvig, *Introduction til Natur- og Folkeretten* (Elektronisk utgave, (2013 [1716]), <https://Holbergsskrifter.no>
- Jarrett, Simon, *Those They Called Idiots: The Idea of the Disabled Mind from 1700 to the Present Day* (Reaktion Books, 2020).
- Justisdepartemente, *Innstilling fra den av Justisdepartementet 11 mai 1922 oppnevnte komité til revisjon av straffeloven* [Straffelovkomiteen]. Første del (Oslo, 1925).
- Justiskomiteen 1901/1902, *Betænkning, afgivet af en af Den norske lægeforening nedsat komité, angaaende de bestemmelser i straffelovudkastet der særlig interesserer lægestanden. Stortingsforhandlinger 1900–1901, 5 d, Dokument nr 1.*
- Looft, Carl, 'Intelligensundersøkelser av skolebarn', *Medicinsk Revue*, 30.10 (1913), 569–585.
- Lovcommissionen av 1828, *Motiver til det i aaret 1832 udgivne Forslag til en Lov for kongeriget Norge angaaende Forbrydelser* (Kristiania, 1835).
- Mülberger, Anette, 'Biographies of a Scientific Subject: The intelligence test', in *Oxford Research Encyclopedias: Psychology* (Oxford University Press, 2022), <https://doi.org/10.1093/acrefore/9780190236557.013.694>
- Mülberger, Anette, Mònica Balltondre and Andrea Graus, 'Aims of Teachers' Psychometry: Intelligence Testing in Barcelona (1920)', *History of Psychology*, 17.3 (2014), 206–222, <https://doi.org/10.1037/a0033339>
- NOU 1974: 17 *Strafferettslig utilregnelighet og strafferettslige særreaksjoner* (Universitetsforlaget, 1974).
- NOU 1990: 5. *Strafferettslige utilregnelighetsregler og særreaksjoner: Straffelovskommisjonens delutredning IV* (Universitetsforlaget, 1990).



- NOU 2014:10: *Skyldevne, sakkyndighet og samfunnsvern* (Departmentenes sikkerhets- og serviceorganisasjon, 2014).
- Porter, Theodore M., *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton University Press, 1995), <https://doi.org/10.1515/9780691210544>
- Prins, Adolphe, *La Défence sociale et les transformations du droit penal* (Misch et Thron, 1910).
- Rosted, Jacob, *Innbydelsesskrift til den offentlige eksamen ved Christiania Kathedralskole* (Jacob Lehmann, 1818).
- Schaanning, Espen, *Kampen om den forbryterske sjel. Kriminal-filosofiske vitenstrekk* (Unipubforlag, 2002).
- Skålevåg, Svein Atle, 'Kjønnsforbrytelser. Sedelighet, seksualitet og strafferett 1880–1930', *Tidsskrift for kjønnsforskning*, 1–2 (2009), 7–27, <https://doi.org/10.18261/issn1891-1781-2009-01-02-02>
- Skålevåg, Svein Atle, *Utilregnelighet: en historie om rett og medisin* (Pax, 2016).
- Skålevåg, Svein Atle, 'A culture of consensus: Organising expertise in Norwegian forensic psychiatry, late nineteenth to early twentieth century', in *Forensic Cultures in Modern Europe*, ed. by Willemijn Ruberg, Lara Bergers, Sara Serrano Martínez, Pauline Dirven (Manchester University Press, 2023).
- Straffelovskommissionen, *Udkast til almindelig borgerlig Straffelov for Kongeriget Norge II. Motiver* (Det steenske forlag, 1896).
- Westrum, Victoria and Linda Gröning, 'Strafferettslig utilregnelighet ved psykisk utviklingshemming—en diskusjon om regelen og praktiseringen av den', *Tidsskrift for strafferett*, 3 (2022), 214–237, <https://doi.org/10.18261/strafferett.22.3.2>

## 4. IQ Testing and Sterilization in Norway, 1930 to 1960

*Per Haave and Jon Røyne Kyllingstad*<sup>1</sup>

---

In 1942, a young woman was sterilized without being informed and without having consented to it, having been tricked into thinking that the operation was meant to relieve menstrual pain. Eleven years later, in 1953, she sued the Norwegian state and the psychiatrist Augusta Rasmussen and demanded 100,000 Norwegian kroner in compensation. The woman had been sterilized by virtue of the Norwegian Sterilization Act of 1934, which legalized the sterilization of persons with intellectual disabilities without their own consent. Intelligence testing played a key role in the implementation of this legal provision, and different perceptions about the woman's intelligence, as measured by IQ, were at the center of the trial. The case led to a public debate about IQ testing, eugenics, and expert power, and contributed to a stricter interpretation of the Sterilization Act and an increasing limitation of the practice of involuntary sterilizations from the late 1950s.

This chapter looks at the implementation of the Norwegian Sterilization Act of 1934, the rise and fall of the notion of a eugenic ideal of preventing mental retardation through sterilizations, and the use of intelligence testing to identify those who could be legally sterilized. Our attention will be directed towards the activities of the national Sterilization Board, a body established to assess all cases of involuntary sterilization in Norway.

---

1 Per Haave was working on this chapter when he suddenly passed away on 18 January 2024 at the age of 66. After Haave's tragic death, Kyllingstad decided to make an effort to complete the work Haave had begun, based on his unfinished draft and numerous conversations with Haave. Large parts of the chapter are an edited version of Haave's manuscript. Some parts are written anew by Kyllingstad after Haave's outline but based on studies of primary sources. These include the sections 'A Lawsuit', 'Disputed IQ Testing', 'The Fall of Eugenic Sterilization', and 'Concluding Remarks'.

## The Norwegian Sterilization Act of 1934

The global history of legal sexual sterilization is rooted in American eugenics. Indiana took the lead by implementing a sterilization act in 1907. In Europe, sterilization acts gained a geographical pivot in the Nordic countries (Denmark: 1929, 1934; Sweden: 1934, 1941; Norway: 1934, 1942; Finland: 1935; Iceland: 1938). Elsewhere in Europe, sterilization laws were introduced in the canton of Vaud in Switzerland (1928), Estonia (1935) and Latvia (1937), while a German sterilization code was enacted shortly after Hitler's rise to power in 1933. After the end of World War Two, American, European, and Nordic sterilization laws remained in force for several decades.

The drafting of the Norwegian Sterilization Act was rooted in a claim from the women's liberation movement in the early 1920s that argued for the castration of sex offenders to be included in the penal code. This claim was, however, rejected by the Penal Code Committee of 1922. Instead, the committee proposed a separate bill regulating sexual surgery (sterilization and castration). The process leading up to the passing of the law in 1934 became strongly influenced by eugenic ideas. While the main intention of the bill was to authorize voluntary sterilization on grounds other than strictly medical—sterilization at that time was punishable under the Penal Code—it also legalized eugenic sterilization of persons with mental deficiency, mental illness, or a serious disease or defect of another kind, not specified.<sup>2</sup> Shortly after the passing of the act, the psychiatrist Johan Lofthus hailed it as the instrument society needed to curb fertility and the greater urge to procreate among the biologically and socially unfit.<sup>3</sup> Lofthus became a long-standing member of the Sterilization Board from 1937 onwards, and as shown by Lefkaditou (chapter 1, this volume), he was the creator of the first proper Norwegian version of the Stanford–Binet test, which came to play an important role in the enactment of involuntary sterilizations.

2 Per Haave, *Sterilisering av tattere 1934–1977. En historisk undersøkelse av lov og praksis* (Oslo: Norges forskningsråd 2000), pp. 88–97.

3 Johan Lofthus, 'Lov om adgang til sterilisering m.v. sanksjonert 1. juni 1934. Foredrag i Opland fylkes lægeforening 25. juni 1934', *Tidsskrift for den norske lægeforening*, 54 (1934), 999–1007 (pp. 1006–1007).

The 1934 Norwegian Sterilization Act was in force until 1 January 1978, excluding a couple of years during the period 1940–1945 when Norway was occupied by Germany.<sup>4</sup> The act established access to sterilization as an individual right. However, this right could be restricted; only persons with full legal rights could apply for sterilization without third-party consent (Section 3–1). In case of *mangelfullt utviklede sjelsevner* [deficient mental capabilities], mental illness, and a young age, the act required third-party consent either from the applicant's natural guardian or a guardian appointed by court (Section 3–2). In case of severe mental illness and 'severely deficient mental capabilities', the person in question was deprived of all legal rights, including the right to apply for sterilization, which had to be requested by a guardian or corresponding authority (Section 4).<sup>5</sup>

The 1934 Sterilization Act did not categorize people in terms of IQ, even though the expert members of the Penal Code Committee of 1922 were familiar with intelligence testing. Instead, the Penal Code Committee proposed the term 'the nine-year mental age'. It was meant to demarcate the line between those who were and those who were not considered competent to consent. The term was introduced by Ragnar Vogt, who clarified that it did not refer to the nine-year "intellectual" age but to intelligence in a wider context, i.e. mental capabilities in a broad sense.<sup>6</sup>

However, the Ministry of Justice and Police found the term 'the nine-year mental age' inapplicable in a law text and replaced it with a new term, adding the word '*særlig*' [severely] to the term '*mangelfullt*

---

4 The Norwegian Nazi regime replaced the 1934 Act with its own sterilization law, 'Act for the protection of the race [Lov nr. 1 til vern om folkeætten]' from 29 December 1942. It remained in force until the end of the German occupation, 7 May 1945. Norway was the only nation outside the Third Reich that introduced an act designed by Nazi authorities. As with the German act, the new Norwegian act authorized sterilization of persons suffering from a range of allegedly hereditary illnesses only, and if necessary, with the help of the police. Other sterilization laws allowed for sterilizations on both social and eugenic grounds. It should also be noted that while the 1934 Norwegian Sterilization Act was in place until 1978, the new law on sterilization, that is still in place today, builds closely on its predecessor, but has no eugenical provisions.

5 *Lov om adgang til sterilisering m.v. av 1. juni 1934.*

6 Haave, *Sterilisering*, p. 95. In the other Nordic countries, the condition to perform sterilization without consent from the person concerned was less restrictive; fourteen years in Finland and twelve years in Denmark and Sweden.

*utviklede sjelsevner*’ [deficient mental capabilities] taken from the Penal Code’s provisions on criminal unaccountability (see Grønning and Skålevåg, chapter 3, this volume).<sup>7</sup> The new wording was meant to cover those who, due to mental deficiency, could not be ascribed a right of self-determination as they were considered unable to reason and understand information, and communicate a choice. In practice, however, it turned out that the term came to cover persons categorized as “imbeciles” and “idiots”, which was generally defined according to IQ scores. Here, “mental age” was translated to IQ, and vice versa, and used interchangeably with “intelligence age”.

While applications under Section 3 had to state an *‘aktverdig’* [decent/worthy] reason, not specifically defined, applications under Section 4 had to state one of three specific reasons, or a combination thereof: social (i.e., when a person, on account of insanity or severely insufficiently developed mental capabilities, was unable to provide for oneself and family and take care of children); eugenic (i.e., when a person was assumed to transmit to the descendants a mental or physical disorder); or prevention of sex crimes (concerned primarily with castration).<sup>8</sup>

All applications had to be accompanied by a medical certificate and, if applicable, written consent from the spouse. The Norwegian Director of Health handled the applications under Section 3–1, while the aforementioned consultative board of experts in sterilization matters (the Sterilization Board) handled the applications under Section 3–2 and Section 4. The Sterilization Board consisted of four members appointed by the government—two medical doctors, a lawyer, and a woman [sic; meant to represent the lay perspective]. It was presided over by the Norwegian Director of Health.

7 Odelstingsproposisjon (Ot.prp.) no. 17 (1934), p. 7.

8 The 1934 Act comprehended both sterilization (elimination of reproduction) and castration (reduction or elimination of sex urge). Section 4 reads: ‘På sinnssyke og personer med særlig mangelfullt utviklede sjelsevner kan seksualinngrep tillates foretatt efter begjæring fra verge eller den i § 6 nevnte kurator, når det ikke er håp om helbredelse eller vesentlig bedring og det er grunn til å anta at vedkommende ikke vil bli i stand til ved eget arbeide å sørge for sig og avkom, eller at en sykkelig sjelstilstand eller en betydelig legemlig mangel vilde bli overført på avkom, eller at han på grunn av en abnorm kjønnsdrift vil begå sedelighetsforbrytelser.’

## Instructions in 1938 and 1950

As already mentioned, the term 'mental age' was a core concept in the Penal Code Committee's proposal (1932), while the Ministry of Justice and Police saw it as inapplicable in a legal text. The term was also absent when the Directorate of Health launched the first instruction to the Sterilization Act in 1938, which instead used the term '*intelligensalder*' [intelligence age]. Persons aged sixteen and older with an intelligence age (IA) of nine years or below were to be regarded as having 'severely deficient mental capabilities'. The same applied to individuals with a particularly low "ethical level" [*etisk nivå*], even if they had an intelligence age above nine years.<sup>9</sup> This meant that the state health bureaucracy extended the possibility of applying Section 4 without change in the legal text.

This actual extension of the scope of the law was a response to growing criticism of the act's strict criteria for imposing sterilization.<sup>10</sup> The criticism came above all from pedagogues working in the special school system, at *åndssvakeskoler* [training schools for "feeble-minded" children] and at *skolehjem* [reformatory boarding schools for children and youth seen as neglected or exhibiting asocial behavior].<sup>11</sup> Additionally, Lofthus, as a member of the Sterilization Board, considered mental defectives with an IA higher than nine years as severely deficient if their 'moral defects' were substantial. In 1938, Lofthus wrote to a psychiatrist that 'significant moral defects may make it necessary to declare individuals with IA > 9 years as severely deficiently developed', adding that this would now be highlighted in the instruction from the Directorate of Health.<sup>12</sup>

A new instruction published in 1950 introduced the term IQ, almost in passing, by comparing the intelligence age scale and the IQ scale: 'In general, people aged 16 and over who have an intelligence age (IA) of 9 years or less (intelligence quotient below 56) will be considered as individuals with severely insufficiently developed mental capabilities.' It further stated that, 'where the IA is 9–12 years old (IQ 56–75), the person

9 'Veiledning i behandling av steriliseringssaker (1938)', Attachment to Haave, *Sterilisering*, pp. 394–396.

10 Some medical doctors and educationalists even wanted to set the limit for Section 4 sterilizations at eleven instead of nine years of intellectual age.

11 Haave, *Sterilisering*, pp. 352–358.

12 'Steriliseringssaker 1938': Til Augusta Rasmussen fra Johan Lofthus (som sinnssykeinspektør) 2.7.38.

concerned will most often be regarded as one with insufficiently developed mental capabilities.' In the new instruction, 'particularly low ethic level' was replaced by 'psychopathic personality traits'. If these traits appeared as dominant symptoms, the person in question could be regarded as one with 'severely insufficiently developed mental capabilities'.<sup>13</sup>

## Eugenic Sterilization of the Feeble-minded

In the interwar period, the question of sterilization of the feeble-minded [*åndssvake*], to use the historical term, gained increasing prominence in medical literature. In the late 1930s, the question became prominent in the politics of intellectual disabilities. It was thoroughly treated by a special group of the Committee of Social Legislation, appointed by the Labor government in 1935 to consider the social security system. In 1945, this special group presented a proposal—almost finished when Germany invaded Norway in 1940—for an act on the care of people with intellectual disabilities. The proposed draft was based on commonly accepted ideas, including the premise that intellectual disabilities, especially *debilitet* [the milder forms of debility], were to a large extent biologically inherited. The Committee's special group also expected the intellectually disabled, except in severe cases, to birth more children than most people, and to cause 'the most difficult social problems' such as asocial and criminal behavior, vagrancy, and prostitution. It was therefore deemed necessary to reduce 'the growing number of defective individuals' by sterilization. However, the special group did not believe that it was possible in the near future to eliminate intellectual deficiency by means of sterilization.<sup>14</sup> Professor Ragnar Vogt was invited to participate in the special group. He was a leading figure among Norwegian psychiatrists, a former key member of the Penal Code Committee, which drafted the Sterilization Act, and had gained expertise in the clinical and social psychiatric understanding of mental deficiency.

13 'Veiledning i behandling av steriliseringssaker (1950)', attachment to Haave, *Sterilisering*, pp. 397–400. A person from the Norwegian Directorate of Health who had participated in the making of the 1938 instruction later remarked that 'particularly low ethical level' had not been considered an apt term, because what they had in mind was the ability to adapt socially.

14 *Sosiallovkomiteen av 1935, Innstilling VIII: Lov om omsorg for åndssvake* (1946), p. 10.

The special group's recommendations on sterilization were rooted in interwar eugenics, in accordance with one of the main motives behind the 1934 Sterilization Act. An expression of how the eugenic solution was continued in post-war politics of mental deficiency is to be found in a study on social patterns in the Nordic countries, published jointly by the Nordic Ministers of Social Affairs in 1953:

[...] in recognition of the great importance of heredity in this field, it is only natural that efforts should be made to prevent an increase in their number. In Finland, Iceland, Norway, and Sweden special legislation, adopted during the 'thirties, authorizes sterilisation when indicated by social, medical or eugenic consideration; the corresponding Danish provisions authorize sterilisation only for social reasons, although the question of heredity is an important factor in arriving at a decision.<sup>15</sup>

In total, between 1934 and 1977, around 44,000 sterilizations (30,000 on women, 14,000 on men) were performed pursuant to the 1934 Sterilization Act, the majority of which were under Section 3–1 from the mid-1960s onwards, when sterilization became more widely used as a means of contraception. The recorded number of eugenic sterilizations, i.e., sterilizations under Section 3–2 and Section 4, reached 1,201 and 922 respectively—2,123 in total. Sterilizations under these two sections peaked in the late 1940s and early 1950s. Thereafter, Section 3–2 surgeries were more slowly reduced than Section 4 surgeries. In the late 1960s, the number of operations on women under Section 3 was still higher than in the 1930s.<sup>16</sup>

Most eugenic sterilizations were performed following the diagnosis of mental deficiency (75 percent of 2,123 sterilizations under Section 3–2 and Section 4 together). This meant that most of these surgeries were executed on persons with a recorded IQ of 75 or below. Even though the code itself did not contain any IQ figures, the handling of the applications was based on the assumption that an IQ of 75 in general constituted the limit between “normality” (or near full normality) and “abnormality”. Those with a recorded IQ above 75 were ascribed full legal rights, while those with a recorded IQ below 56 were deprived of all legal rights—including the right to apply for sterilization—on the grounds

---

15 George R. Nelson and Aune Mäkinen-Ollinen, *Freedom and Welfare: Social Patterns in the Northern Countries of Europe* (København, 1953), p. 371.

16 Haave, *Sterilisering*, p. 154.



that they were considered unable to reason and deliberate, understand information, and communicate a choice. For those with a recorded IQ between 56 and 75, the right to apply for sterilization was, however, an illusory form of autonomy. Based on the requirements set by the law and the situation of the persons in question, it is rather doubtful that the applications qualify as autonomous choices. Thus, it is reasonable to refer to sterilizations under Section 3–2 as forced sterilizations, although these surgeries according to the letter of the law were voluntary.<sup>17</sup>

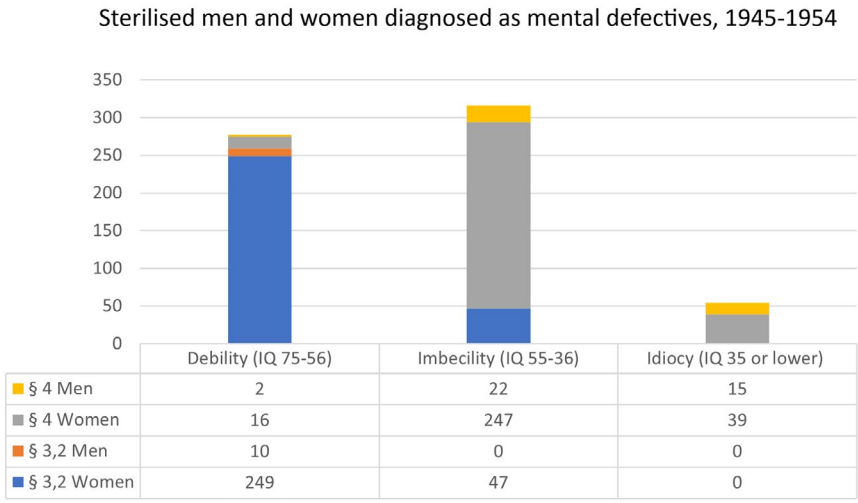


Fig. 4.1 Table showing an overview of the number of men and women diagnosed as “mental defectives”, 1945–1954, compiled by the authors. For source information see Karl Evang, ‘Sterilisering etter lov av 1. juni 1934 om adgang til sterilisering m.v.’ Bilag til *Tidsskrift for Den norske lægeforening*, nr. 2, 1956: 16 (Table 2), 31 (Table 10), 39 (Table 12).

The Sterilization Board made use of established categories of mental deficiency—idiocy, imbecility, and debility—and related IQ values that were generally accepted in Norwegian medical discourse. Thus, most Section 3–2 surgeries were performed on the diagnosis of debility (IQ 75–56), while most Section 4 surgeries were performed on the diagnosis of imbecility (IQ 55–36). That relatively few sterilizations were performed on the diagnosis of idiocy (IQ 35 or lower) was partly owing to the assumption that idiocy was more often due to diseases in early

17 Haave, *Sterilisering*, pp. 154, 182–198.

childhood or injuries in early age or during pregnancy than biological inheritance; and partly since “idiots” rarely gave birth to children. In all categories, women outnumbered men. Figure 4.1 illustrates the distribution of sterilized mental defectives by regulations, gender, and intelligence categories in a certain period (1945–1954).<sup>18</sup>

How does the number of eugenic sterilizations relate to the estimated population of mental defectives? In the 1930s as well as in the 1960s, one percent of the population at any time was estimated to be intellectually deficient.<sup>19</sup> In Norway, this estimate included individuals with an alleged IQ of 75 or below. It was further assumed that the frequency of mental deficiency was higher among men than women, approximately sixty against forty percent. A systematic policy of selective sterilization of those born in the period 1900–1959 would have placed 14,000 mentally defective women and 21,000 mentally defective men in the risk group. In practice, ten percent of the women and less than one percent of the men in the risk group were sterilized.<sup>20</sup>

There was no break with the practice of eugenic sterilization in the immediate postwar years. As we have seen, the number of eugenic surgeries in Norway peaked in the late 1940s. The same was the case in Sweden and Denmark.<sup>21</sup> In 1955, Karl Evang—the long-standing Norwegian General Director of Health (1938–1972) who presided over the Sterilization Board—still believed in inheritance as a significant cause of mental deficiency, and hence advocated for more extensive eugenic sterilization. Evang acknowledged the deterrent effect of Nazi practices on eugenic sterilization, but he commented hopefully that there had now been time ‘to work this off and discuss the problems on an objective basis’.<sup>22</sup>

18 According to Evang (*Sterilisering etter lov av 1. juni 1934 om adgang til sterilisering m.v., Bilag til Tidsskrift for Den norske lægeforening*, nr. 2 (1956), p. 16), the distribution of IQ categories had to be taken with certain caveats, as a psychiatrist was not always involved in the assessment of the individual’s mental state. This is also in accordance with Haave’s readings of the medical records in the so-called sterilization archive (Haave, *Sterilisering*, pp. 208–221).

19 K. Evang, *Sterilisering etter lov*, p. 37.

20 Haave, *Sterilisering*, s. 154–155.

21 Mattias Tydén, ‘The Scandinavian States: Reformed Eugenics Applied’, in *The Oxford Handbook of the History of Eugenics* (Oxford: Oxford University Press, 2010).

22 K. Evang, *Sterilisering etter lov*.



Fig. 4.2 Karl Evang (1902–1981), Norwegian General Director of Health 1938–1972, he was involved in the establishment and leadership of the World Health Organization. During the interwar years, he became known to the Norwegian public as a radical socialist, an eager advocate for sexual education, family planning and the legalization of abortion. In the pamphlet *Rasepolitikk og reaksjon* [Racial Politics and Reaction] (1934), he strongly criticized racial thinking and a racist and classist style of eugenics, while advocating a non-racial and class-neutral style of eugenics, as part of a scientifically organized socialist society.

Photographer Rigmor Dahl Delphin/Oslo Museum, CC0 1.0.

Despite the relatively modest sterilization of mental defectives, Karl Evang was convinced, even in 1960 (reprinted in 1970), that the surgical solution had ‘had some influence in reducing inherited feeble-mindedness.’<sup>23</sup> Evang, who had hoped for a more extensive eugenic

23 Karl Evang and Dorothy Burton Skårda, *Health Services in Norway*, 2nd ed. (Oslo, 1960), p. 101 and 3rd edition, p. 126.

sterilization, had two main explanations for the relatively low number of surgeries. On the one hand, he considered the division of labor and clientele between the fields of education and medicine as artificial and as an obstacle for a eugenically-oriented policy towards those with an IQ above 56. According to the 1949 Act on financing of institutions for the mentally defective, 55–56 was an IQ threshold between those deemed uneducable and therefore belonging to the field of medicine, and those deemed educable and therefore belonging to the education system.<sup>24</sup> On the other hand, he held that the idea of *senmodning ettermodning* [late maturation], which was gaining ground in the postwar society, had blurred the etiological distinction between inheritance and environmental factors, and caused a more restrictive attitude towards sterilization. “Late maturation” meant that an individual could increase their intelligence at a later point than the average, due to changed environmental conditions or a slower development of the brain. Evang held, however, that this was irrelevant for the eugenic justification for sterilizations, as hereditary factors remained unaffected by late maturation.<sup>25</sup>

## The Sterilization Board’s Processing of Applications

The process of determining a person’s mental state involved a complicated mixture of factors. In an interview in 1953, Evang underscored that the Sterilization Board always sought to consider the entire situation when assessing sterilization, and that intelligence testing was an important, but far from the only element in such an assessment. In what way were IQ scores made relevant in the sterilization board’s assessments of individual cases?

In Section 4 cases, the important point was to decide whether the person was an “imbecile”, and thus had severely insufficiently developed mental capabilities, and therefore could be sterilized without

---

24 The Ministry of Church and Education Affairs had the political-administrative responsibility for those with an IQ of 56 and above. The Ministry of Social Affairs for those with an IQ of 55 and below. The medics, like Karl Evang, wanted the responsibility of the entire system of care for the mentally defective. The twofold politics and administration were confirmed by the 1949 Act on financing of institutions for the mentally defective.

25 K. Evang, *Sterilisering etter lov*, p. 38.

their own consent. In Section 3.2 cases, the decision was if the person was a “debile” and thus had insufficiently—but not severely so—developed mental capabilities, and could be sterilized on their own request, provided consent from a guardian.

As mentioned, an intelligence age of nine and an IQ of 56 were regarded as the threshold between imbecility and debility, with an IQ of 75 as the upper limit for debility. In the total assessment of a sterilization case, these cut-off points were, however, not seen as decisive. In line with the 1938 instruction, Section 4 sterilization was sometimes applied to people with an IQ above 56, based on the argument that the person’s ‘ethical level’ was particularly low, or if the person had ‘psychopathic personality traits’ as formulated in the 1950 instruction. Cases like this were often subjected to discussions and doubt among the board members, especially in the 1950s and later. According to a talk by psychiatrist and Sterilization Board member Nic Waal in 1955, it was important to assess to what extent the ‘psychopathic condition’ or low ethical level was transient and curable, or whether it was so stable that the person could be considered to have severely insufficiently developed mental capabilities.<sup>26</sup>

Shifting views on the heritability and stability of intelligence also had implications for the assessment of sterilization cases, and these implications differed depending on the sections in question. Eugenics had been a key motivation behind the Act, and eugenic arguments were sometimes taken into consideration. Some applications for sterilization were rejected on the grounds that the person did not have a hereditary condition, and that a sterilization could not be recommended on eugenic grounds. In other cases, a sterilization was recommended to prevent continuation of a hereditary condition. Eugenic arguments were, however, not decisive in most sterilization cases. In Section 4 cases, in particular, the main question was about the person’s ability to take responsibility for a child. A person deemed unable to decide about their own sterilization was also considered incompetent to become a parent. In these cases, the eugenic argument was not decisive for the sterilization decision, and it was therefore not necessary to know whether or not the condition could be inherited.<sup>27</sup>

---

26 Talk by Nic Waal referred in *Nordisk Medicin*, 54, no. 49 (1955), p. 1830.

27 Haave, *Sterilisering*, pp. 215–221.

Eugenic arguments and the question about heritability were more relevant in Section 3.2 cases. Evang and others held that genetics was more often involved in debility than in imbecility, and that debile persons often gave birth to many children. But even in Section 3.2 cases, the eugenic argument was often less important than the question about the applicant's ability to take care of a child, including the potential mother's social, economic, and familial conditions. Evang, Waal and other members of the board voiced frustration over the lack of resources to assist these women with contraceptive advice and alternative contraceptive methods.<sup>28</sup>

While the question about the heritability of intelligence was not necessarily important in Section 4 cases, the question about the stability of intelligence throughout a person's life was highly relevant. It was problematic to forcibly sterilize an imbecile person if it was likely that this person's intelligence could, later in life, rise to a level where they were competent enough to make informed decisions about sterilization or parenthood. This issue became an important topic in the Sterilization Board's assessment, and contributed to a more restrictive attitude towards forced sterilizations during the 1950s, along with the increasing support for theories of 'late maturation', and of 'pseudo-debility' and 'pseudo-imbecility'; an acknowledgement that weak intellectual abilities could be the result of unfortunate environmental circumstances and could be amended.

## A Lawsuit

The complexity of the sterilization cases as well as the varying quality of the assessments done in these cases were put on the public agenda in the early 1950s. A woman who was sterilized at the age of nineteen in 1942 pursuant to the 1934 Sterilization Act brought a lawsuit against the Ministry of Social Affairs and the psychiatrist who wrote the expert report that the Sterilization Board had based their decision on. The case was brought all the way to the Supreme Court in 1956. The woman claimed that the sterilization had been processed and approved on grounds of gross negligence and was therefore unlawful. The expert report, written by psychiatrist Augusta Rasmussen, claimed that the applicant suffered from hereditary, permanent, and severe mental deficiencies (i.e., severely

---

28 K. Evang, *Sterilisering etter lov*; Haave, *Sterilisering*.

insufficiently developed mental capabilities). The intelligence test, carried out by Rasmussen, had resulted in an intelligence age of seven years and six months, corresponding to an IQ of 47, well below the threshold for sterilization in compliance with Section 4 of the Sterilization Act. The woman was a legal minor at the time, with only her father providing consent. Rasmussen had considered it unnecessary to obtain consent from the woman's mother, since she had a history of mental illness.<sup>29</sup>



Fig. 4.3 Augusta Rasmussen (1895–1979) was, in 1933, the second woman in Norway to become a specialist in psychiatry. The same year she succeeded Johan Lofthus in the position of school psychiatrist in Oslo (1933–1965). In 1932 she published the award-winning study 'The impact of sexual assaults on children under 14 years of age for the development of mental disorders and character anomalies'. Here, she claimed that most of the children she had examined had suffered no bodily or mental injury from the sexual assault. In 1947, she published a study based on the IQ testing of 310 Norwegian women who had had sexual relationships with German occupation soldiers, and claimed to have found that nearly all of them were more or less intellectually disabled. In retrospect, her research has faced strong criticism, but in her time, she was a recognized professional and was often used as an expert witness in the courtroom. Photographer unknown/Arbeiderbladet/Arbeiderbevegelsens archive and library. All rights reserved.

29 Haave, *Sterilisering*, pp. 364–370.

The young woman was never informed about the sterilization and was under the impression that she was undergoing surgery to relieve menstrual pain. Before the Court, she challenged the grounds for the medical evaluation of her intelligence. Moreover, the applicant highlighted procedural errors, as most members of the Sterilization Board—including the Norwegian Director of Health—had been absent when her application was approved. She claimed 100,000 Norwegian Kroner in compensation, a substantial sum at the time.<sup>30</sup>

The City Court (the Court of First Instance) found no error in psychiatrist Rasmussen's assessment of the applicant's intelligence, and while acknowledging the absence of committee members in the decision-making process as a procedural mistake, they found no liable error in the actions of the authorities or the individual psychiatrist. After an appeal, and another loss in the second instance with a 5–2 division among the judges, she tried to take the case to the Supreme Court, where it was rejected. The final judgment of the Court of Appeal confirmed the ruling of the first instance, even though it found multiple procedural flaws: particularly that the applicant had not consented to, or been duly informed about her sterilization, and that the Norwegian Director of Health had not participated in the Sterilization Board's decision-making process. Along similar lines as the first instance, the Court of Appeal pragmatically ruled that the outcome would likely not have differed with the presence of the Norwegian Director of Health; in doing so, it rejected the applicant's claims.

The case attracted significant media attention and catapulted the issue of coercive sterilization and IQ testing into the public discourse. The social-liberal newspaper *Dagbladet* even covered the sterilized woman's legal expenses for the appeal when the government—based on the argument that she was doomed to lose the case—denied her free legal aid.<sup>31</sup> The public debate went beyond the specific court case and highlighted a number of issues related to IQ testing and its role in society, such as the quality of the tests; the relationship between testing and discretionary judgement; the role and limits of psychiatric, psychological, pedagogical, and legal expertise; and last but not least, the question about the heritability and stability of intelligence. The

---

30 Haave, *Sterilisering*, pp. 364–370.

31 *Dagbladet*, 7 May 1954.



heated debate tells us something about prevailing beliefs and attitudes among key experts and in the wider public about “intelligence” and IQ, and its role in society at that point in time.

A key topic was the discrepancy between the IQ scores obtained by Augusta Rasmussen when she examined the woman in 1939 and 1942, and the higher IQ scores obtained by another expert who re-tested the woman in 1951. The 1942 psychiatric report stated that the woman’s low intelligence could not be expected to increase in the future, and that the woman would not be able to provide for herself (and a potential child) through her own work. A decade later, the woman was, in fact, providing for herself and partly for her mother through work, and the new test put her above the IQ threshold for imbecility. In spite of this, both the municipal court and the majority in the Court of Appeal found that psychiatrist Rasmussen was not to blame for her failed prognosis, since she had written her expert report at a time when there had been a general agreement among the experts that a person’s intelligence could not improve after the age of sixteen.<sup>32</sup>

This argument was most likely based on the testimony of one of the expert witnesses, the renowned psychiatrist Hans Henrik Dedichen, who claimed that ‘the woman was a victim of the slow development of science’. According to him, British and American researchers had now found out that environmental change could create ‘an astonishing development’ in people ‘with very low intelligence’, and if this had been known in 1942, it was unlikely that Rasmussen would have proposed to sterilize the woman. When Rasmussen wrote her report, however, Lofthus’ book *Intelligensmåling* (1931) was the only work on ‘intelligence measurements in this country’, and, according to Dedichen, it represented a ‘more absolute view’ on ‘the constant size of the intelligence quotient’. The quality of IQ tests had also increased since 1942, and he admitted that the new test in 1951 ‘gave a better picture’ of the woman’s mental state. Dedichen concluded that Augusta Rasmussen was thus not to blame, since she had acted in good faith based on the dominant scientific understanding at that point in time.<sup>33</sup>

Dedichen was opposed by the psychiatrist Ole B. Munch, a

---

32 *Dagbladet*, 27 November 1953.

33 *Dagbladet*, 18 November 1953.

senior physician at *Statens hjem for åndssvake* [the State Home for the feebleminded], state consultant on mental retardation issues, and responsible for a national register of the feebleminded. Munch maintained that Dedichen's line of argument was in conflict with the spirit of the law—on the one hand, Dedichen recognized that the woman was currently too intelligent to be sterilized; and on the other hand, he claimed that it had been justifiable to sterilize the woman in 1942 since she had been judged an imbecile at that time. If a transient condition were to fall under the Sterilization Act's concept of 'severely insufficiently developed mental capabilities', very many of us could 'fall into the well and be exposed to a medical intervention'.<sup>34</sup>

In addition to the low IQ score, Rasmussen's conclusion in the 1942 report was based on claims about the woman's assumed "low ethical level". The case exemplifies in an interesting way what an expert assessment of "ethical level" could look like. Rasmussen explained to the court that she had based this assessment on the fact that the girl, according to her father, 'was babbling, making things up, and adding' in a way that a person at her age should be 'done with'. The girl was also unusually interested in boys. This was the reason her father had requested sterilization. Rasmussen had also concluded that the woman would never be able to provide for herself and a potential child. When the defense lawyer pointed out that the woman was actually providing for herself through a job at a café, Rasmussen argued that it was now much easier to get a job than it had been before the war. As an example of this, she mentioned that she knew of a vagrant girl who had gotten a job in a cafe and that, compared to the vagrant girl, 'the girl here in court is a beauty'.<sup>35</sup>

## Disputed IQ Testing

According to Ole B. Munch, the City Court's ruling was based on a 'vulgar use of the intelligence quotient'. It placed too much emphasis on the IQ figure in isolation, and disregarded the rest of the psychological assessment. He also pointed to the lacking precision of IQ tests. Based

---

34 *Dagbladet*, 18 November 1953.

35 *Dagbladet*, 8 March 1956.

on his own discretionary judgement, he claimed that even the re-test in 1951 had produced a far too low estimate of her intelligence.<sup>36</sup>

While Munch criticized bad test practice and the incorrect use of test results, others went further and attacked the very practice of measuring intelligence.<sup>37</sup> The non-expert Roar Hauglid, who was a well-known public figure, art historian, and from 1958, national antiquarian,<sup>38</sup> described the sterilization of the woman as an ‘abhorrent and outrageous’ case that had caused a public outrage. The problem was not that the psychiatrist had done a bad job, but that ‘the intelligence testers’ in general ‘had managed to convince themselves and others that they [...] can measure human souls’. He hoped that the case would help to reveal this *sjelelige kvakksalveri* [quackery of the soul], which was going on ‘under the guise of the name of science’.<sup>39</sup>

While Augusta Rasmussen’s long experience as a school psychiatrist was used in the courtroom as an argument for her competence, Hauglid turned this argument on its head and used the sterilization scandal as a starting point to criticize the ‘unscientific’ IQ testing of schoolchildren in order to ‘dupe’ parents and transfer children to special schools. He wanted ‘the psychologists’ out of the school, leaving it up to teachers and parents to decide what was in the children’s best interest.<sup>40</sup> This led to a public debate with the psychologists Dagny Oftedal, Arvid Ås, and Åsa Gruda Skard.

The school psychologist Arvid Ås claimed that the outrage was understandable and that a lot of ‘psychological quackery’ was going on in Norway. The problem was not the testing, but that people without professional psychological training—including teachers and psychiatrists—whose testing practices had major consequences, such as declarations of insanity, forensic psychiatric considerations, and sterilizations, carried out irresponsible testing. Responsible testing required both broad psychological knowledge and specific expertise in psychometrics and statistics, and, according to Ås, only

36 *Dagbladet*, 1 December 1953.

37 *Dagbladet*, 19 November 1953, 26 November 1953 and 10 December 1953.

38 *Lillehammer Tilskuer*, 23 November 1953; *Dalane Tidende*, 23 November 1953; *Vestfold Fremtid*, 9 December 1953.

39 *Dagbladet*, 19 November 1953.

40 *Dagbladet*, 26 November 1953.

university-educated pedagogues and psychologists had this expertise in Norway.<sup>41</sup> Åsa Gruda Skard agreed with Ås that all testing should be carried out by university-trained pedagogues and psychologists, that it was important to have good cooperation and a proper division of labor between the teacher, the doctor and the psychologist, and that psychologists in general were critical to the lack of caution in the testing of intelligence.<sup>42</sup>

The sterilization case debate included strong expressions of distrust towards intelligence tests and psychological and psychiatric expertise. It also included, as seen above, leading psychologists who promoted their own profession's jurisdiction over test practices and tried to delegitimize the allegedly incompetent use of IQ tests by other professions, including psychiatrists. This is interesting to note, because not only had psychiatrists taken the lead in the introduction of IQ testing and the development of IQ tests in Norway, but also because the debate was about the Sterilization Act, which regulated a medical practice under the Health Directorate, and that it was a psychiatrist who was sued for misconduct.

A recurring theme in the debate was the question of the heredity and constancy of intelligence. The discussions rested on the same premise as the judgment of the Court of Appeal, namely that the idea of intelligence as a hereditary and constant factor had been disproved, and replaced by the realization that intelligence could be influenced by environmental conditions and that a person's intelligence could undergo a "late maturation".<sup>43</sup>

The heritability question was also a key issue when a biology student, Anton Brøgger, directed sharp criticism against the Sterilization Act in November 1953. He maintained that terribly little was known about the inheritance of psychological traits, and doubted that it could be proven that 'mentally ill-equipped people have equally ill-equipped children'. He maintained that there was no agreement on what intelligence was, that intelligence measurements were riddled with errors, and that it was outrageous that they were used to decide on the right to have children.

---

41 *Dagbladet*, 26 November 1953.

42 *Dagbladet*, 28 November 1953.

43 *Dagbladet*, 30 November 1953.

He also put forward ethical arguments: To allow involuntary sterilization on the grounds that a person was 'exceptionally ungifted' epitomized an ideal of creating a 'humanity [...] of mentally well-equipped people', and could be compared to the Nazi's extermination of allegedly inferior Jews.<sup>44</sup>

In 1959, Brøgger became a research fellow at the Department of Human Genetics in the Faculty of Medicine at the University of Oslo.<sup>45</sup> Among other things, this department worked with a register of Norwegians with potentially hereditary diseases, including the intellectually disabled, and eugenic ideas were part of the background for the establishment of this register. Brøgger later became head of research at the Department of Genetics at the Radium Hospital in Oslo. In this context, it is interesting that he, as a young man, had so strongly rejected the idea of eugenic sterilizations and the scientific and ethical foundation of the Sterilization Act.<sup>46</sup>

In contrast, Health Director Karl Evang—the man who had the highest administrative responsibility for implementing the Sterilization Act—saw nothing in the sterilization case to suggest that the law should be revised. As already mentioned, in the mid-1950s Evang was still advocating eugenic sterilization, and in a *Dagbladet* interview, he claimed that the Sterilization Act was 'a straightforward and simple law [...] unbureaucratic and practical' with 'flexible' provisions. He pointed out that 'perhaps the most important purpose' of the act 'has fallen somewhat into the background' in the debate about the sterilization case, namely that it was a 'eugenics law'.<sup>47</sup>

---

44 *Dagbladet*, 30 November 1953.

45 *Årsberetning Universitetet i Oslo*, 1959–60.

46 In 1975, Brøgger translated and edited the Norwegian edition of American geneticist L.C. Dunn's book *Heredity and Evolution in Human Populations*, which discusses eugenics. "Old eugenics" is rejected, as based on outdated scientific beliefs, and the complexity of heredity and environment are underscored. Among other things, Dunn argues that intellectual disability is relative to societal conditions and more people will become intellectually disabled when society makes increasing demands on individuals' ability to think. Dunn advocates a form of eugenics, but in the form of (contraceptive) guidance for (potential) parents who are carriers of hereditary diseases. Sterilization is not mentioned, and certainly not forced sterilization.

47 *Dagbladet*, 2 December 1953.



Fig. 4.4 Nic Waal (1905–1960) was a psychiatrist, one of Norway's first psychoanalysts, a pioneer in child psychiatry, and, from 1951, founder and director of the Nic Waals *Institutt for barne-og ungdomspsykiatri* [The Nic Waal Institute of Child and Adolescent Psychiatry], which is still a leading institution in its field in Norway. Like Karl Evang, she was a member of the left-wing radical student group *Mot Dag* in her youth. In 1943, she was imprisoned by the German occupying authorities for her participation in the resistance movement. Photo: Arbeiderbladet 1958/Arbeiderbevegelsens arkiv og bibliotek, CC BY NC ND 4.0.

When the Court of Appeal's judgment was available, all documents in the case were sent to the members of the Sterilization Council. Two members commented on the matter. The most substantial commentaries were given by psychiatrist Nic Waal. She found the medical information incomplete and partly misleading. Considering the woman's difficult background, her ability to work and the fact that she had no children out of wedlock, Waal found it 'highly questionable' that the case had been brought forward at all. She found the method of misleading the woman 'completely reprehensible'. Waal also found it highly 'worrying' that

the case had been sent to 'a Nazi sterilization board', and that this had been done shortly after the current board had been dismissed, among other things, due to principled disagreement on the interpretation of the Sterilization Act. Waal, who had herself been imprisoned during the war for her participation in the resistance movement, claimed that 'everyone' knew 'that the Norwegian Nazis with the director of medicine leaned towards the German conception of sterilization. [Rasmussen] therefore had to know that the case would not be judged by an unbiased sterilization board'. Waal had no doubt that the girl had been misjudged. She was also surprised that the members of the current sterilization board had not previously been allowed to comment on the case, since the members had been discussing similar matters since the 1930s. She herself had then, and later, repeatedly pointed out that any IQ score was insufficient to assess a personality, and that statements about immorality very often were highly subjective.<sup>48</sup>

Even if the Ministry of Social Affairs and the psychiatrist were acquitted, and even if Evang claimed that the case gave no reason to reconsider the Sterilization Act, the case did affect the administrative practice in a restrictive direction. It clearly appears in the discussion of concrete cases that the General Directorate of Health and the Sterilization Board feared new lawsuits.<sup>49</sup> However, this change was not caused by the case in court alone.

## An Increasingly Restrictive Practice

During the 1950s, the Sterilization Board carried out an increasingly restrictive practice, particularly in the case of sterilization under Section 4. The more restrictive administrative practice was not only due to the court case and the ensuing public debate. It was also due to the consideration of principles.

In the first three years of the decade, nearly ten percent of the applications were rejected, compared with 7–8 percent in the period 1945–1950. In the period 1954–1956, 28.6 percent of these applications were rejected, and in the succeeding three-year period, it was 38.1 percent. The main reason for rejection was that the Board did not find

---

48 Haave, *Sterilisering*, p. 368.

49 Haave, *Sterilisering*, p. 364–370.

it sufficiently documented that the person concerned was incompetent to consent to sterilization. In the case of sterilizations based on Section 3-2, the number of rejected applications was more stable and ranged between six and nine percent in the period from 1948 to 1962. In these cases, the main reason for rejection was that the Board doubted that the applications reflected a voluntary choice from the person concerned.<sup>50</sup>

In other words, the professional judgment forming the basis of the applications was increasingly evaluated from the point of view of a strict interpretation of the formal criteria stated in the law. This also meant that the Norwegian Directorate of Health applied the conditions of using Section 4 stated in the law in a more restricted manner. The new instruction (1950) emphasized that Section 4 could not be used if the client 'understood' or 'could be made to understand the significance of the operation'. IQ was meant to function as the dividing line in cases where consent hinged upon the person's presumed ability to comprehend the consequences of sterilization. In 1952, a committee for preparing proposals for a revision of the legislation on the care of the insane from 1848 proposed that this amplification should be included as an additional condition in Section 4 of the Sterilization Act. This amendment of the act was passed in 1961, to the effect that Section 4 could only be used if the patient did not have 'the capacity to make a decision concerning the operation'. Since the practice was already guided by the instruction of 1950, the new wording did not actually lead to a changed practice, since the important change of practice had already taken place.

The changed practice did not necessarily mean that IQ tests were given less importance in sterilization cases. Perhaps, on the contrary, they even gained increased importance. When the question of the person's ability to understand and make decisions about the sterilization question came into focus—and when the individual's rights were increasingly emphasized, with the threat of legal consequences if experts and authorities made incorrect assessments—the question of intelligence level became more important and the room for discretionary assessments of 'ethical level' less important.

---

50 Haave, *Sterilisering*, p. 240–245.



## The Fall of Eugenic Sterilization

At the same time as the requirements for consent were being tightened, opposition to eugenic sterilizations increased. Even though the Director of Health was far from being alone in his advocacy of eugenic sterilizations of the feeble-minded in the mid-1950s, the idea met increasing scientific and normative critique and suffered a loss of support and professional legitimacy in the years that followed.

Critics argued that hereditary defects are mainly inherited recessively, so that most carriers of the hereditary factor cannot be identified, that there are often complex combinations of various hereditary and environmental factors behind a mental retardation, and that sterilizations would have little effect on the prevalence of these factors. Also, the notion that IQ tests are reliable and precise, and that they measure an innate quality was criticized, along with the idea that a person's level of intelligence is stable throughout life. Particularly important was the latter idea, as it implied that an IQ score obtained at one point in life does not necessarily provide a prognosis for the development of intelligence. Many of these arguments had already been put forward in the interwar period, but they now gained greater impact, in parallel with the emergence of new normative attitudes towards people with intellectual disabilities.

In the 1930s, it had been common to describe intellectually disabled people as inferior. Among both educators and medical doctors, it was common to see sterilization presented as a key method to prevent mental retardation in future generations, alongside segregated and institutionalized care. With the ambitious post-war development of the welfare state, these principles were put into practice with increasing efficiency. The number of involuntary sterilizations peaked in the late 1940s and early 1950s alongside the early phase of a period of rapid growth of institutions for the feeble-minded. Right after the war, there were only a few hundred places in institutions for the intellectually disabled. In 1969 this had grown to about 5,500 places spread over 100 institutions. The vast majority of these were at boarding institutions where the intellectually disabled lived twenty-four hours a day.<sup>51</sup> Paralleling the increasing institutionalization, however, a gradual change in attitudes

---

51 *Norges offentlige utredninger*, 25 (1973), 14.

towards people with intellectual disabilities was coming about.<sup>52</sup> The new attitude was strongly expressed by an expert committee which the Ministry of Social Affairs set up in 1963 to assess the work that was being done for 'physically, mentally and socially handicapped' children and youths, and propose future directions of this work.<sup>53</sup>

The committee maintained that heredity played a role in both physical and psychological "defects" and diseases, including 'intelligence defects', and that genetics could provide information which was relevant to an individual's own decisions about whether or not to have children.<sup>54</sup> The committee underscored, however, that heredity was far less important than previously thought, and unanimously rejected the eugenic idea that prevention of intellectual disability could be based on sterilization. This idea was both 'offensive for humane reasons' and inefficient, since most cases of hereditary intellectual disability were due to recessive inheritance, which meant that most carriers were not intellectually disabled themselves.

Instead, the committee wanted to place emphasis on the prevention of environmentally caused developmental disabilities, and on treatment and education of the intellectually disabled, to develop their individual potential and quality of life. The committee pointed out that the mentally retarded had previously only received practical training, because it was assumed that they would not benefit from theoretical education, and argued that new research had demonstrated the efficiency of adapted teaching. It highlighted the fact that teaching positions had now been established at all major institutions for the intellectually disabled. The future lay in treatment, training and prevention of environmentally caused developmental disabilities, not in prevention through sterilization.<sup>55</sup>

The committee report called for the "normalization" of people with disabilities and underscored their equal rights to healthcare and education, but did not challenge the principle of institutionalized care. However, the 1960s also saw the rise of new organizations and activists who fought for the rights of the disabled, and from this movement,

---

52 Anne-Lise Seip, *Veiene Til Velferdsstaten: Norsk Sosialpolitikk 1920–75* (Oslo: Gyldendal, 1994), p. 248.

53 Bjørn Skau, *Innstilling om omsorgen for handikapte* (Oslo: Sosialdepartementet 1965).

54 Skau, *Innstilling om*, p. 10.

55 Skau, *Innstilling om*, p. 15.

sharp criticism was also eventually directed against the very principles of segregation. In the mid-1970s, this criticism gained political traction, and the Norwegian Parliament adopted a comprehensive reform, which led to the phasing out of the large state institutions and the integration of the mentally disabled into local communities.

## Concluding Remarks

Eugenic arguments formed an important background for the passing and purpose of the Sterilization Act, in that it legalized involuntary sterilization of the feeble-minded and mentally ill with the intention of reducing the incidence of hereditary mental weakness in the population. But the act also had two other main purposes: to prevent people who were unable to care for a child from having children, and to legalize voluntary sterilization for individuals with a worthy reason.

Although the act did not mention IQ at all, IQ testing came to play a significant role in assessing an individual's rights, or lack thereof, to sterilization, with or without consent. But the IQ test was never considered conclusive. The concept *mangelfullt utviklede sjelsevner* [deficient mental capabilities] was not identical to a concept of permanently impaired intelligence, and the IQ test was part of an overall psychiatric assessment where 'ethical level' was also an important criterion. This gave a lot of space for the psychiatrists' professional judgment.

Although the eugenic goal of limiting the incidence of mental retardation through sterilizations had great support among psychiatrists and educators in the first two decades of the act's existence, the majority of non-voluntary sterilizations were justified by a person's inability to take care of a child. From the mid-1950s, eugenic justifications became even less common. At the same time, the criteria for carrying out sterilizations without consent were tightened. The key question was whether the person concerned had the ability to understand what a sterilization entailed, and what consequences it would have. Paralleling this shift in the interpretation and application of the Act, the dominating view on IQ and intelligence was modified. The idea that an IQ test measures an inherited and immutable characteristic was increasingly replaced by the view that intelligence, as measured by an IQ test, is a product of both heredity and environment, and that a person's level of

intelligence is not necessarily given once and for all.

The story that has been told in this chapter may perhaps appear to be about scientific progress—the revelation, through new research in the 1950s, that the pre-war scientific consensus about heritable and immutable intelligence was wrong; and the consequent changes in attitude and practice towards the intellectually disabled. This is how the story was narrated by many psychologists and psychiatrists who participated in the debates about sterilization, eugenics, and the care for the intellectually disabled in the late 1950s and early 1960s. But the problem with this narrative is that it corresponds with reality only to a limited extent. Even before the war, there were several critics who put forward well-founded scientific arguments against both exaggerated notions of the heritability of intellectual traits and against the efficiency of eugenic solutions. It may be that what happened towards the end of the 1950s was not primarily the result of new scientific insights leading to new attitudes and practices—but rather that the breakthrough of new attitudes and ideals led to the scientific burden of proof being reversed, and that it became easier to gain traction for scientific arguments against genetic determinism and eugenic solutions.

## Bibliography

- Haave, Per, *Sterilisering av tatere 1934–1977: En historisk undersøkelse av lov og praksis* (Oslo: Norges forskningsråd, 2000).
- Lofthus, Johan, 'Lov om adgang til sterilisering m.v. sanksjonert 1. juni 1934. Foredrag i Opland fylkes lægeforening 25. juni 1934', *Tidsskrift for den norske lægeforening*, 54 (1934), 999–1007.
- Nelson, George R. and Aune Mäkinen-Ollinen, *Freedom and Welfare: Social Patterns in the Northern Countries of Europe* (København, 1953).
- Evang, Karl, 'Sterilisering etter lov av 1. juni 1934 om adgang til sterilisering m.v.', *Bilag til Tidsskrift for Den norske lægeforening*, nr. 2 (1956), 16.
- Evang, Karl and Dorothy Burton Skårdal, *Health Services in Norway*, 2nd ed. (Oslo, 1960), and 3rd ed. (Oslo, 1970).
- Tydén, Mattias, 'The Scandinavian States: Reformed Eugenics Applied' in *The Oxford Handbook of the History of Eugenics* (Oxford: Oxford University Press, 2010), <https://doi.org/10.1093/oxfordhb/9780195373141.013.0022>
- Årsberetning Universitetet i Oslo 1959–60* (Oslo: University of Oslo, 1960).

Seip, Anne-Lise, *Veiene Til Velferdsstaten: Norsk Sosialpolitikk 1920–75* (Oslo: Gyldendal, 1994).

Skau, Bjørn, *Innstilling om omsorgen for handikapte* (Oslo: Sosialdepartementet 1965).

Waal, Nic., *Untitled talk, Nordisk Medicin*, 54.49 (1955), 1830.

*Norges offentlige utredninger* 1973 (25): Omsorg for psykisk utviklingshemmede. Målsetting og retningslinjer (Oslo: Universitetsforlaget/Sosialdepartementet 1973).

*Sosiallovkomiteen av 1935. Innstilling VIII: Lov om omsorg for åndssvake* (1946).

*Odelstingsproposisjon (Ot.prp.) no. 17 (1934). Lov om adgang til sterilisering.*

# 5. Tools of the Trade: Psychotechnics, IQ Testing and the Making of the Psychological Profession in Norway, 1925–1947

*Isak Lønne Emberland*

---

While the IQ test is arguably one of the best known, widely disseminated, and historically significant types of psychological tests, it is far from the only one. IQ tests have historically often been used in conjunction with other types of psychological assessments, and they continue to be used today. During the first half of the twentieth century, a common name for psychological testing and applied psychology in general was ‘psychotechnics’.<sup>1</sup> Since the psychological profession today is identified with clinical activity in Norway, the history of the profession has commonly been written with a focus on the clinical aspect.<sup>2</sup> The strong—and somewhat unique—connection between academic psychology and psychoanalysis is also a distinguishing feature of the early years of Norwegian psychology.<sup>3</sup> But, as I will show, the stages of early professionalization were also marked and defined by a testing practice, psychotechnics, separate both from psychoanalysis and the clinical in general.

- 
- 1 Although the term “psychotechnics” has had differing meanings, where the terms have ranged from describing applied psychology in general, to a more specific term used to denote, for instance, “industrial psychology”, it was broadly known and used.
  - 2 Andreas Høstmælingen, ‘Hvordan Norsk Psykologi Ble En Klinisk Profesjon’, *Tidsskrift for Norsk Psykologforening*, 59.4 (2022), 248–67; Håvard Friis Nilsen, ‘Psykologene: Universitetsfaget Som Ble En Klinisk Profesjon’, in *Profesjonshistorier* (Pax, 2014), pp. 450–70; Bergljot Baklien, *Psykologprofesjonen Og Dens Omgivelser, Utredninger Om Forskning Og Høyere Utdanning*, 6 (NAFs utredningsinstitutt: Norges almenvitenskapelige forskningsråd, 1976).
  - 3 Håvard Friis Nilsen, *Resistance in Therapy and War: Psychoanalysis in Norway 1920–1945 (Doktorgradsavhandling)* (Universitetet i Oslo, 2016); Håvard Friis Nilsen, *Du må ikke sove: Wilhelm Reich og psykoanalysen i Norge*, 2. opplag (Aschehoug, 2023).

This chapter explores how the early years of psychotechnic testing practice shaped the psychological profession in Norway. Part of this story includes examples of psychotechnical test construction—as with the ‘goldsmith test’, created by Helga Eng at the Oslo Psychotechnical Institute at the Oslo Vocational School in the late 1920s. Especially in the early days of applied psychology, there were few “ready-made” solutions. Thus, a certain number of tests had to be assembled, and sometimes entirely new tests had to be constructed to deal with the task or problem at hand. Probing for specific psychological qualities in a particular epistemological, material, and institutional setting necessitated a certain degree of improvisation on the psychologist’s part, alongside cooperation with the client’s needs and desires. As I will illustrate in this chapter, the process of developing a psychological test battery was thus often marked by professional discretion.

Overall, this chapter empirically traces how the task of psychotechnical testing led to a professionalization of psychology in Norway, with the establishment of a role and a way of working distinct from disciplinary, academic work. In doing so, the chapter also sheds light on the way psychotechnics operated in a context beyond countries commonly considered to be the epicenters of the psychotechnical movement in Europe, namely Germany, France, and the UK.

## A Note on the History of Psychotechnics and its Relationship to Intelligence Testing

Psychotechnics was one of the drivers in the early professionalization of both European and American psychology. Psychotechnical testing aided psychologists in carving out a separate jurisdiction from other professions, such as medicine, centered around questions and problems pertaining to efficiency, rationalization, and optimalization of the workforce.<sup>4</sup> The profession of psychology in America especially gained legitimacy through its use of intelligence tests. While this is also true for Europe, here it was through the more heterogeneous practice of aptitude testing of workers in industry, in an attempt to raise production. This had already been

---

4 Ulfried Geuter, *The Professionalization of Psychology in Nazi Germany* (Cambridge: Cambridge University Press, 1992); John Carson, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750–1940* (Princeton, NJ: Princeton University Press, 2007); Nikolas Rose, *Governing the Soul: Shaping of the Private Self*, 2. utg. (London: Free Association Books, 1999); Nikolas Rose, *Inventing Our Selves: Psychology, Power and Personhood* (Cambridge University Press, 1998).

attempted through Taylorism (scientific management), whose principles had been applied to industrial workplaces on both sides of the Atlantic during the early 1900s. But whereas scientific management mainly focused on the optimization of the physical spaces and processes in factories and workshops, psychotechnics bore with it the promise of both selecting and optimizing the working “soul”.<sup>5</sup> One of the main fields of psychotechnical intervention was the task of selecting apt workers for different jobs. “The right man for the right place” was the often-repeated motto.

The term “psychotechnics” is often attributed to the German psychologist and philosopher William Stern, who delineated practical psychology from theoretical psychology. Stern is also credited as the originator of the concept of the intelligence quotient, or IQ. Although his specific understanding of both IQ and psychotechnics would not be widely accepted during the twentieth century, the dual origin of these influential terms points to their intertwined historical and conceptual relationship.

Based on the history of the development and application of psychotechnical methods, one could argue that the professionalization of psychology in this period was marked by an interplay between scientific and pragmatic epistemologies, and between psychologists and non-academic clients.<sup>6</sup> The tools, techniques, and ideas of psychotechnical testing were, from the start, in dialogue with the constraints placed on them by the client, in the form of a task or a work assignment. As with, for instance, the implementation of IQ tests in the American military during World War One, the tests themselves, as John Carson argues, were shaped by the interplay between the army officials’ ‘presupposed ideas’ and the psychologists’ scientific ideals. Here, interpersonal relationships and psychologists’ charisma, or lack thereof, became important, and arguably deciding factors, for which tests were employed.<sup>7</sup>

Although the institutionalization and employment of psychotechnical

5 Anson Rabinbach, *The Human Motor: Energy, Fatigue and the Origins of Modernity* (Berkeley: University of California Press, 1990); Anson Rabinbach, *The Eclipse of the Utopias of Labor* (New York: Fordham University Press, 2018); David Meskill, *Optimizing the German Workforce: Labor Administration from Bismarck to the Economic Miracle* (New York: Berghahn Books, 2010); David Meskill, ‘Psychological Testing and the German Labor Market, 1925 to 1956’, *History of Psychology* 18, nr. 4 (2015), 353–66; Charles S. Maier, ‘Between Taylorism and Technocracy: European Ideologies and the Vision of Industrial Productivity in the 1920s’, *Journal of Contemporary History*, 5, nr. 2 (1970), 27–61.

6 Pieter J. van Strien, ‘Early Applied Psychology Between Essentialism and Pragmatism: The Dynamics of Theory, Tools, and Clients’, *History of Psychology*, 1, nr. 3 (1998), 205–34; John Carson, ‘Army Alpha, Army Brass, and the Search for Army Intelligence’, *Isis*, 84 (5 April 2011), <https://doi.org/10.1086/356463>.

7 Carson, ‘Army Alpha, Army Brass, and the Search for Army Intelligence’.



methods in industry and education are often referred to as consequential steps towards the professionalization of psychology, the broader aspects of what psychotechnics was are often overlooked or confounded in the general historiographic literature. This is not necessarily due to lack of scholarly diligence, but is rather caused by the widely different ways psychotechnics was understood in its time. The term was very much in flux in its heyday in interwar Europe. In the US, the term 'psychotechnics' went out of fashion after World War One and was, in many cases, supplanted with 'industrial psychology' or 'organizational psychology'.

To better understand the Norwegian case, it is helpful to picture psychotechnics as referring to two distinct, albeit often intertwined, aspects. Firstly, psychotechnics refers to a broad techno-scientific movement with its epicenter in continental Europe. Although carried forward by scientific ideas, this movement also comprised political, cultural, and economic ambitions that surpassed strictly academic pursuits. The word "psychotechnics" could denote an array of loosely defined, but evocative ideas pertaining to the use of psychological science for the betterment of society.

Secondly, psychotechnics has a more limited meaning, denoting the application of certain methods for physical and psychological measurement. These methods were taken from the brass instrument tradition of experimental psychology, but whereas psychological experiments were used to produce scientific knowledge, psychotechnical testing was used mainly to rate people according to their ability for a specific type of work. This could then be used to assess people's abilities for employment or enrolment at a workplace or a school. This version of psychotechnics had a more limited scope.

I use these two "types" of psychotechnics more as heuristic tools than analytical categories—as ways of organizing the sometimes chaotic use of a loosely defined term. However, they have some explanatory power when it comes to the question of how psychotechnics circulated and was "claimed" by certain professional groups.

The concept of intelligence played an important role in both these versions of psychotechnics. Intelligence testing was one of many ways of probing suitability for different types of work, and thus could be used to harness the psychological forces for the benefit of society and

the individual. In the context of the concrete practice of psychotechnical testing, such as in the case I will explore later, intelligence tests often played a role as a part of a test battery. In both these cases, psychotechnics contributed to disseminate intelligence tests, and embedded them in discursive and institutional frameworks. Thus, both the psychotechnical movement, and the more limited psychotechnical practice, can be understood as historical foundations for the apparent staying power of intelligence testing. In Norway, this is connected to how psychotechnics led to the professionalization of psychology and established a certain psychological practice.



Fig. 5.1 Harald Krabbe Schjelderup (1895–1974), a Norwegian psychologist and one of the authors of the white paper establishing psychology as a stream of professional education at the University of Oslo in 1948. Photographer unknown/ Museum of University History, KHM, University of Oslo. All rights reserved.

## Testing and the Psychological Profession

Psychology arguably did not become a legally recognized profession in Norway before 1948, when the Norwegian parliament adopted a temporary law establishing an *embetseksamen* [professional exam] in psychology at the University of Oslo. Most scholars working in the field of the history and/or theory of the professions agree that professions cultivate and make use of expert knowledge to the benefit of society.<sup>8</sup> This knowledge is practiced with a high degree of autonomy and discretion, and, generally speaking, can be protected through gatekeeping mechanisms, such as academic institutions.<sup>9</sup> A profession is given (or claims, depending on whom you ask) a jurisdiction over certain types of societal areas of intervention, sometimes at the expense of another profession.<sup>10</sup> Historically, I argue, a constitutive factor for a profession is whether its core epistemic assumptions, the normative grounds on which the profession intervenes in society, and the knowledge-based tools it employs to further these ends are recognized by society at large. This means that social and cultural legitimacy are especially essential for professions, including those that intervene with human subjects such as medicine and psychology.

Taking these definitions into account, the establishment of the professional exam in 1948 marks the beginning of the psychological profession proper in Norway. Other scholars have pointed out that

---

8 Andrew Abbott, *The System of Professions: An Essay on the Division of Expert Labor* (Chicago: University of Chicago Press, 1988); Eliot Freidson, *Professionalism: The Third Logic*, Nachdr. (Cambridge: Polity Press, 2011); Harold L. Wilensky, 'The Professionalization of Everyone?', *American Journal of Sociology*, 70, nr. 2 (September 1964), 137–58, <https://doi.org/10.1086/223790>; Thomas Brante, *Den professionella logiken: Hur vetenskap och praktik förenas i det moderna kunskapssamhället* (Stockholm: Liber, 2014).

9 Lisa Wallander and Anders Molander, 'Disentangling Professional Discretion: A Conceptual and Methodological Approach', 2014, <https://doi.org/10.7577/pp.808>

10 The theory of professions as a system where each competes over jurisdiction is explored in Abbott, *The System of Professions: An Essay on the Division of Expert Labor*; a similar type of structure, pertaining to science, scientific disciplines and science and non-science, is explored in Thomas F. Gieryn, 'Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists', *American Sociological Review* 48, nr. 6 (1983), 781–95, <https://doi.org/10.2307/2095325>; Thomas F. Gieryn, *Cultural Boundaries of Science: Credibility on the Line* (Chicago: University of Chicago Press, 1999).

societal and professional foundations for this new professional role were laid during the interwar years.<sup>11</sup> As mentioned, scholars have mainly studied the historical and intellectual strains that came to form psychology as a clinical profession, i.e., oriented towards providing therapy as a part of healthcare. However, the *stortingsmelding* [white paper] that became the foundation of the 1948 law was distinctly marked by an altogether different ambition for the psychologist and their role in society.

In the white paper—written by psychologists Harald Schjelderup, Åse Gruda Skard, and Johannes Clausen—psychology is largely legitimized by the claimed societal need for categorization, sorting, and differentiation of individuals. Schools, the military, and, most importantly in this context, the labor market, were underscored by the authors as areas and institutions in need of psychological expertise. What was proposed was especially the training of psychologists that could work in the field of applied psychology, namely psychotechnics.<sup>12</sup> This idea resonated with the political project of the welfare state in the postwar years. In 1947, the same year the proposal was written, the Norwegian prime minister, Einar Gerhardsen, spoke at the first Nordic psychological congress. Gerhardsen belonged to the Norwegian Labor Party, which, after World War Two, established a political hegemony that lasted until the 1980s. ‘In all countries,’ he claimed, ‘the societal development has made it necessary for society itself, the state, the municipality with all its branches, to intervene, in a larger degree than before, into the everyday occupation of the individual citizen, both in his work and outside it’.<sup>13</sup> To ensure ‘harmonious societal development’, there was a need for psychologists that could test, assess, and categorize

11 Håvard Friis Nilsen, ‘Psykologene: Universitetsfaget som ble en klinisk profesjon’, in *Profesjonshistorier* (Oslo: Pax, 2014), pp. 450–70; Andreas Høstmælingen, ‘Hvordan norsk psykologi ble en klinisk profesjon’, *Tidsskrift for Norsk psykologforening*, 59, nr. 4 (2022), 248–67; Fredrik W. Thue, ‘In Quest of a Democratic Social Order: The Americanization of Norwegian Social Scholarship 1918–1970’ (Oslo: Universitetet i Oslo, 2006); Fredrik W. Thue and Kim G. Helsvig, *1945–1975: Den store transformasjonen, Universitetet i Oslo 1811–2011 5* (Oslo: Unipub, 2011).

12 Harald Schjelderup, Åse Gruda Skard, and Johannes Clausen, ‘Ot. prp. nr. 98 (1947): Lov om eksamen i psykologi ved Universitet i Oslo’, *Stortingsforhandlinger* (ib.utg.), vol. 91, nr. 3 (1947).

13 Emil Østlyngen (red.), *Nordisk psykologmøte i Oslo 1947* (Oslo: Gyldendal, 1948), p. 16. (All quotes from Norwegian sources are translated by the author.)

people for the right jobs.<sup>14</sup>

The psychological test, and in particular tests that could help in sorting the workforce according to aptitude, thus played an important role in the legitimization and justification of the psychological profession. The test, and not only the intelligence test, was a distinctive tool that made the societal work of psychology possible.

### Psychotechnics in Norway Prior to 1925

The authors of the white paper, and prime minister Gerhardsen, based their claims on an already existing practice, stemming from the interwar years, which they wanted to expand and consolidate in the hands of trained psychologists. Psychotechnics became properly institutionalized in Norway in 1925, but as early as 1914, there were attempts to create institutions supporting psychotechnical practice. As in many other European nations, psychotechnics were perceived by Norwegian actors as a way of rationalizing industry, with the desired result of increasing profits. In addition, some actors (but not all) regarded psychotechnics as an instrument for the emancipation of the individual. The idea was that if everybody had the opportunity to be tested for their innate attributes and skills, and thereby be selected for a job best suited to their psychological makeup, this would lead to a greater increase in both individual happiness and societal benefit. Thus, there was little to no perceived conflict between individual and societal needs, although the two were underscored differently by different actors.

For psychologists, the possibility of an institution with psychotechnical tasks meant a potentially greater influence on society and more jobs for psychologists. Psychotechnics was seen as a necessary next step in going from a purely theoretical science to an applied one. Echoing one of the most well-known theorists and advocates of psychotechnics, Hugo Münsterberg, the Norwegian professor of psychology Anathon Aall wrote in 1919 that 'through psychology, philosophy works in collaboration with the natural sciences and with

---

14 Østlyngen (red.), p. 17.

practical life, as it manifests itself in commerce and industry'.<sup>15</sup>

Aall is commonly regarded as one of the founding fathers of the psychological discipline in Norway, and made efforts to demarcate psychology as an experimental, empirically founded science.<sup>16</sup> As illustrated above, Aall regarded psychotechnics as an important way of making psychology relevant for society, something he often wrote about in public during the 1910s and 1920s. He would stress the need for the use of scientific methods in 'practical life', and saw psychotechnics as a potential benefit to other scientific disciplines, such as pedagogy, law, and psychiatry.<sup>17</sup> Other Norwegian psychologists, such as Martin Luther Reymert and Thorleif Grüner-Hegge, held several lectures during the early 1920s about psychotechnics.<sup>18</sup> They also corresponded with actors such as the Norwegian Industrial Association, where they put forward applications for funding of psychotechnical research with the promise of future benefits for the association.<sup>19</sup> These research programs never manifested in something concrete, but they demonstrate how psychotechnical research and methods from an early point in time were marketed and partially funded by actors in industry and commerce outside academia. As discussed later, this dissemination of psychotechnical ideas, literature, and methods outside academia led to the interest in and demand for psychotechnics, and thus the recruitment of psychologists in psychotechnical positions.

---

15 *Bergens Annonse Tidende* 1919.02.25, *Bergens Annonse Tidende* 1919, *Bergens Annonse Tidende*, 1, <https://urn.nb.no/URN:NBN:no>

16 The early interwar years were marked by an increasingly institutional, as well as epistemic, separation of psychology from philosophy. Whereas psychology traditionally had been a sub-discipline of philosophy, the new, experimental psychology sought closer alignment with empirical science, in particular the natural sciences. This is especially visible in their distancing from the Hegelian philosophy that had dominated the university. With this new orientation also came ambitions to put psychology to use.

17 When discussing the plans for a new university in Bergen, Aall argued, on these grounds, that psychology, and psychotechnics in particular, should occupy a central place in the curriculum. *Bergens Annonse Tidende* 1919.02.25, 1.

18 Reymert was one of the first who attempted to standardize IQ tests in Norway, before he went on to work in the US. See Lefkadtou, chapter 1, this volume.

19 *Aftenposten* 1922.01.21, *Aftenposten* 1922, *Aftenposten*, 1922, [https://urn.nb.no/URN:NBN:no-nb\\_digavis\\_aftenposten\\_null\\_null\\_19220121\\_63\\_40\\_3](https://urn.nb.no/URN:NBN:no-nb_digavis_aftenposten_null_null_19220121_63_40_3); 'Pr. Arkiv nr. 636 Norges industriforbund. Riksarkivet. 'Psykoteknikk' (sak nr. 224)', u.å.



Fig. 5.2 Anathon Aall (1867–1943), Norwegian philosopher and psychologist. Established experimental psychology at the University of Oslo. Photographer Fritz Möller, 1904/National Library of Norway, PDM 1.0 (Public domain).

Another example that illustrates both the professional and disciplinary heterogeneity, as well as the far-reaching societal and economic ambitions often associated with psychotechnics, can be found in a 1921 proposal put forward to the former prime minister of Norway,

Gunnar Knudsen from *Venstre* [the Liberal Party], by economist Wilhelm Keilhau (1888–1954), pharmacist Klaus Hansen (1895–1971), and aforementioned psychologist Harald Schjelderup.<sup>20</sup> The authors proposed a loosely defined but ambitious plan for a *Centralinstitut for arbeidsteknikk* [Central Institute for the Science of Work] based at the University of Oslo. With a formulation that resembled the style and ethos of European and American psychotechnicians and labor theorists from the same period, they claimed that while the forces of nature had increasingly come under the direct control of science and society, the forces of the human psyche were still a largely untapped resource.<sup>21</sup> Finding a better—that is to say more scientific—and more efficient way of isolating and directing this hitherto unused force, they argued, would be transformative for the competitiveness of the Norwegian economy against the more industrialized nations of the continent. They proposed that this harnessing of the “psychological” force be carried out on a national scale.<sup>22</sup>

Although psychotechnics was just one of many types of ‘science of labor’ discussed in the proposal, it was given a sizable place. The plan was to implement it throughout the country by means of psychotechnical offices. This was probably based on the German model,<sup>23</sup> where,

20 It is unclear whether the authors of the proposal, Keilhau, Clausen, and Schjelderup, intended to approach Knudsen while he was still in office. In the letter sent to him along with the plan, Keilhau refers to him as prime minister. By 1921, however, Knudsen had left office.

21 Examples of this type of rhetoric in the Nordic countries can be found in Alfred Lehmann, *Størst mulig udbytte af legemligt og aandeligt arbejde* (bind 2), bd. 2 (København: J. Frimodts Forlag, 1919); Lehmann; a historical analysis and discussion of the European science of work that preceded the psychotechnical movement in the twentieth century can be found in Anson Rabinbach, *The Human Motor: Energy, Fatigue and the Origins of Modernity* (Berkeley: University of California Press, 1990); Anson Rabinbach, *The Eclipse of the Utopias of Labor* (New York: Fordham University Press, 2018); a discussion of the American case, with special emphasis on Hugo Münsterberg can be found in Matthew Jr. Hale, *Human Science and Social Order: Hugo Münsterberg and the Origins of Applied Psychology* (Philadelphia: Temple University Press, 1980); Jeremy Todd Blatter, ‘The Psychotechnics of Everyday Life: Hugo Münsterberg and the Politics of Applied Psychology, 1887–1917’ (Doctoral dissertation, Harvard University, 2014), <https://dash.harvard.edu/handle/1/12274284>

22 ‘Pr. Arkiv nr. 626 Gunnar Knudsen. Riksarkivet. ‘Politisk virksomhet, utenrikspolitiske saker’’, u.å., Riksarkivet.

23 For a thorough look at the way psychotechnics was used in Weimar Germany, see Meskill, ‘Psychological Testing and the German Labor Market, 1925 to 1956’;



according to historian David Meskill, the labor offices—often the sites for psychotechnics—‘blanketed the country’ during the early 1920s.<sup>24</sup> The individual worker could be tested and receive their ‘psychological roll sheet’ for evaluation by current or future employers.<sup>25</sup> This was a clear and transparent document stating their special competences and latent or actualized psychological potential, both for their own and potential employers’ benefit. Although the institute never materialized, the proposal sheds light on how Norwegian actors from different disciplinary and professional backgrounds collaborated across boundaries with regards to psychotechnics, and how psychologists, through psychotechnics, came to fashion a new, more directly involved societal role as scientists of a nascent profession. It also shows how actors with differing political ideologies could find common ground through scientific labor management and psychotechnics: during the Nazi occupation of Norway, Klaus Hansen would collaborate and actively support the occupiers, while Harald Schjelderup would be imprisoned for his resistance against the Nazi regime.

Interdisciplinarity, collaboration with other professions, and a perceived political neutrality, I argue, are especially important for the psychologists and how they came to be identified with the new, applied science of psychotechnics. As a new disciplinary science, still trying to find its institutional foothold and demarcate its epistemic boundaries vis-à-vis philosophy, biology, and medicine, psychology gained what can be called a professional identity throughout the 1920s in Norway. It was “professional” in the sense that psychologists gradually came to claim some sort of jurisdiction over specific societal problems, namely efficiency and rationality in the labor market. This was, of course, not the only area of expertise singled out by psychologists in this period, but the case can be made that it was the most visible to other disciplines

---

David Meskill, *Optimizing the German Workforce: Labor Administration from Bismarck to the Economic Miracle* (New York: Berghahn Books, 2010); Ulfried Geuter, *The Professionalization of Psychology in Nazi Germany* (Cambridge: Cambridge University Press, 1992).

24 David Meskill, ‘Psychological Testing and the German Labor Market, 1925 to 1956’, *History of Psychology* 18.4 (2015), 354.

25 ‘Pr. Arkiv nr. 626 Gunnar Knudsen. Riksarkivet. ‘Politisk virksomhet, utenrikspolitiske saker’.

and parts of society.

The picture that emerges from these and other examples portrays psychotechnics as a type of knowledge technology that, from the beginning, was disseminated not only among psychologists, but also among actors in industry, politics, and education.<sup>26</sup> The interest in the application of psychotechnics was, however, not the same as the technical skills necessary to make use of it. Thus the “movement” of psychotechnics helped to create the societal demand for a professional group capable of utilizing psychotechnical tools. The psychologist as a handler of psychotechnics in part emerged as a response to this demand. In this early phase, psychologists were not the only group that made a claim to this type of tool. In Germany, for instance, a rather heated dispute broke out between psychologists and engineers over the ownership of psychotechnics and who should be allowed to use it.<sup>27</sup> No such jurisdictional dispute can be found in Norway in the 1920s. What we can see, however, is that both the malleable character of psychotechnics and its association with engineering and adjacent types of knowledge and ideas (such as Taylorism) led to demands for psychological knowledge within new sectors, such as the labor market. Additionally, the interaction between psychologists and non-psychologists became important in shaping the first psychotechnical institution in Norway, namely the Oslo Psychotechnical Institute at *Oslo fag- og forskoler* [the Oslo Vocational School].<sup>28</sup>

---

26 Carl Bonnevie, ‘Praktisk og eksperimentell psykologi i erhverv og skole’, *Sociale meddelelser*, 8.2 (1920), 553–58.

27 Anson Rabinbach, *The Human Motor: Energy, Fatigue and the Origins of Modernity* (Berkeley: University of California Press, 1990); Anson Rabinbach, *The Eclipse of the Utopias of Labor* (New York: Fordham University Press, 2018).

28 In Norwegian, this name is plural, ‘*Oslo fag- og forskoler*’, denoting that the schools that were put under joint administration still operated with a degree of autonomy. To simplify, I will, however, use the singular form in this chapter, since matters pertaining to the Psychotechnical Institute concerned the school (or schools) as a whole.



Fig. 5.3 Helga Eng working on psychotechnical tests, photo taken some time between 1925–1935. Photographer unknown/Oslo Museum, CC0 1.0.

## Establishing the Psychotechnical Institute in Oslo

Psychotechnics entered the Oslo Vocational School as an answer to problems arising from shifting demographics, an increase in specialization in the labor market, and a growing concern within the school's administration that their established practices of admission were increasingly inadequate. The school itself was established in 1921 by merging several smaller schools for artisan, industrial, and skilled work, whereby the different lines were put under a joint administration. In a retrospective on the school's history, office manager Christian Rømme wrote that the reason for the establishment of the school was the industrialization that took place in Norway after 1900. New tools, specializations, and working methods necessitated this

educational reform.<sup>29</sup> The rapid transformation of the labor market led to a need for formalized vocational training, in addition to the traditional apprenticeship under the tutelage of a master. Increasing bureaucratization of the artisan enterprise also resulted in a growing distance between the apprentices and the master, whose time was increasingly occupied by office work.<sup>30</sup> Therefore, the tutorship that had traditionally fallen to the master now had to be partially taken over by formalized education.<sup>31</sup>

Thus, we have several implied causes for the use of psychotechnics. Rømming, and by extension the school's administration, perceived several problems with their existing practices for admission, pertaining to larger historical, societal, and normative causes. By normative causes, I here refer to the fact that they regarded the existing processes of admission as lacking, and sometimes outright unjust. Were school grades really the best indicators of suitability when probing for qualifications and future success in industrial and artisan work? According to Rømming, it was felt that those who aspired to work with tools should be identified by different, or at least supplementary, criteria.

Psychotechnics was, to a large degree, marked by the ambition of probing, scoring, and evaluating attributes not normally reflected in school grades. Although aligned in the promise of revealing the raw potentiality of the individual, we can already see here a certain discrepancy between the goals of psychotechnical testing and intelligence testing: While intelligence was used as a measurement that, by and large, had academic potential within the school as its primary focus, psychotechnical testing tried to probe aptitudes that were traditionally suppressed or devalued in normal schooling.<sup>32</sup> Many Norwegian actors involved with psychotechnics during the interwar years explicitly stated that the Norwegian schooling system did not prepare its students for

---

29 Christian Rømming, *Oslo Fag- og Forskoler for Håndverk og Industri 1921–1946: Et tilbakeblikk* (Oslo: Særtrykk, 1946), p. 3.

30 Rømming, p. 3–4.

31 In the US, intelligence testing was used to differentiate and sort students into schooling in manual or industrial labor. This practice came after the end of the Jim Crow laws and became a substitute for (or a covert continuation of) racial segregation. Mainly black students were assigned to a tracked, secondary school programme. Porter, in this volume, therefore, speaks of 'deskilling'. See Porter, chapter 13, this volume.

32 The appeal to meritocracy and the overcoming of traditional hierarchies was still similar in both psychotechnics testing and intelligence testing.

employment in sectors such as industry, artisanal labor, or agriculture. A school preparing students for such employment, the Oslo Vocational School, thus naturally became one of the first places to make use of psychotechnical testing.

The school's director, Johan Grøn Lund (1878–1951) was an important instigator for the adoption of psychotechnics at the Vocational School. He was educated as a machinist and had a background as the foreman of several technical associations. He was therefore not in any particular contact with psychologists, but instead with the technical professions. Sticking with the aforementioned “typology”, he belongs in the “movement” category since he did not perform the technical work of testing himself. It is plausible that it was through the technical milieu that he first encountered psychotechnics. The first experiments with psychotechnics at the Vocational School, for instance, were conducted by a philologue and a typologist, and were direct applications of existing German tests for typographers. The results, however, were deemed unsatisfactory due to inaccuracy.<sup>33</sup> The solution was to find a person with better knowledge of experimental psychology and a higher diligence when dealing with test scores.

## Helga Eng and the Test for Goldsmiths

The psychologist and pedagogue Helga Eng (1875–1966) was thus employed as the first head of the institute in 1925. This was one of the very first salaried positions for a psychologist outside academia in Norway. Eng had a background in both experimental psychology, which she had studied in part in Germany, and as a schoolteacher.<sup>34</sup> Her research object of choice throughout her career—which in 1938 would lead her to become the first professor of pedagogy at the University of Oslo—was children and their development.<sup>35</sup> Norwegian psychotechnics thus, due in part to being first located at the Vocational School and being overseen

33 Elisabeth Lønnå, *Helga Eng: Psykolog Og Pedagog i Barnets Århundre* (Fagbokforlaget, 2002), p. 153.

34 Lønnå, *Helga Eng: Psykolog Og Pedagog i Barnets Århundre*, pp. 154–56; Kim Helsvig, *Pedagogikkens Grenser: Kampen Om Norsk Pedagogikk Ved Pedagogisk Forskningsinstitutt 1938–1980* (Abstrakt Forlag, 2005).

35 Helga Eng, *Abstrakte Begreper i Barnets Tanke Og Tale. Psykologiske Undersøkelser Paa Grundlag Av Iagttagelse Og Ekspe-Rimenter Med Skolebarn* (Aschehough, 1912).

by Eng, came to be marked by a special attention to children and youths, and like its Danish counterpart, became more closely affiliated with schooling.<sup>36</sup> In contrast, psychotechnics in, for instance, Germany was in large part characterized by close collaboration with industry and the testing of adults.<sup>37</sup> Through Eng in particular, psychotechnics in Norway additionally came to be associated with the pedagogical reform movement. As Helsvig shows (chapter 12, this volume), Eng's international network included key figures in the progressive pedagogical movement. The goal of pedagogy, Eng had stated, was to align itself with a progressive societal development, and secure individual happiness and engagement in a rapidly changing world.<sup>38</sup>

She was not an "orthodox" psychotechnician, neither in methods nor in outlook.<sup>39</sup> Furthermore, although her scientific methods were rooted in empirical, experimental psychology, she has not been regarded as a positivist. In her work, she also employed hermeneutical and literary theory, with a special emphasis on the connection between expression and artistic sensibilities, and psychological development. Her first psychotechnical test shows how her interests in these connections shaped her practice at the institute.

The first original test that Helga Eng developed for the Vocational School was the test for goldsmiths. Earlier tests were primarily imported, and sometimes modified, batteries developed by German psychotechnicians; they were, so to speak, ready-made solutions, well tested and understood as directly applicable to the Norwegian context. There existed, however,

---

36 Peter Triantafillou and Alfonso Moreira, 'Modern Templates of Happiness: Performing Spiritualism and Psychotechnics in Denmark', *History of the Human Sciences*, 18.2 (2005), pp. 87–109.

37 David Meskill, *Optimizing the German Workforce: Labor Administration from Bismarck to the Economic Miracle* (Berghahn Books, 2010); Anson Rabinbach, *The Eclipse of the Utopias of Labor* (Fordham University Press, 2018); Horst Gundlach, 'Psychology as Science and as Discipline: The Case of Germany', *Physis; Rivista Internazionale Di Storia Della Scienza*, 2006, pp. 61–89.

38 See Helsvig, chapter 12, this volume.

39 Hardly anyone can be said to have been "orthodox" or "traditional". Psychotechnicians throughout Europe were usually very heterogeneous in their practices, and one finds a large variation in methods throughout the interwar years. When I say that Eng was not orthodox, I refer to the fact that she placed a greater emphasis on "qualitative" methods than many other psychotechnics working at psychotechnical institutes in, for instance, Germany. Although psychometrics was still the cornerstone of her practice, she also deviated from this in many of her tests, and expressed reservations on the usefulness of these methods in public.

no ready-made test for goldsmiths, and thus Eng had to develop one from scratch.<sup>40</sup> Her work, as she describes it, sheds light on what went into the construction of the test, as well as the role the intelligence test (in this case called a ‘maturity test’) played in the context of the battery.

The battery itself had two modules: one with tests performed individually, and one with group tests, performed in a classroom. Whereas the individual tests were practical or psychomotorical in nature, meant to gauge the subject’s aptitude for goldsmithing, the group test consisted mainly of an intelligence test to measure the general mental aptitude of the youths, and determine their “mental age”. Eng had earlier worked on a Norwegian revision of the Binet–Simon test, developing it at a school in Oslo before she apparently canceled the project in favor of the position at the Psychotechnical Institute.<sup>41</sup> While the Binet–Simon test was conducted individually, the one used for goldsmiths was a group test of her own making, adapted ‘by the author on the basis of the best American tests of its kind, and in our experience with it, it seems to fulfil its purpose’.<sup>42</sup> In total, there were ten individual subtests or tasks in the practically oriented test. The written account was based on experiments with applicants, and thus had its results published. The first of these, described in Eng’s written account, were performed on fourteen adolescent applicants to the school, ranging from the ages of fourteen to eighteen. The first two tests were adapted from earlier German tests for metal workers, although some modifications were made to reflect the more fine-tuned work of goldsmiths. In the first test, the adolescents were to hammer a rectangular lead rod into a quadratic shape, and then hammer a four-pointed tip. In the second test, they were to bend a metal thread into the prescribed shape of a four-leaf clover. They had nine minutes in total, and the time they spent was recorded by the psychologist.<sup>43</sup>

After this, the adolescents underwent a tremometer test, meant to probe their fine musculature and their aptitude for detailed work. Eng wrote that since goldsmiths worked mainly with their hands instead of

40 Helga Eng, *Psykoteknisk prøve for elever ved Oslo guldsmedforskele* (1926), p. 1, [https://urn.nb.no/URN:NBN:no-nb\\_digibok\\_2017112148091](https://urn.nb.no/URN:NBN:no-nb_digibok_2017112148091)

41 Elisabeth Lønnå, *Helga Eng: Psykolog og pedagog i barnets århundre* (Oslo: Fagbokforlaget, 2002), p. 151.

42 Helga Eng, *Psykoteknisk prøve for elever ved Oslo guldsmedforskele* (1926), p. 4 [https://urn.nb.no/URN:NBN:no-nb\\_digibok\\_2017112148091](https://urn.nb.no/URN:NBN:no-nb_digibok_2017112148091)

43 Helga Eng, *Psykoteknisk prøve for elever ved Oslo guldsmedforskele* (1926), pp. 1–2, [https://urn.nb.no/URN:NBN:no-nb\\_digibok\\_2017112148091](https://urn.nb.no/URN:NBN:no-nb_digibok_2017112148091)

machines, the standard metallurgical test did not adequately account for the skill required. The tremometer was thus an inclusion based on her own judgement of the work. The test involved a square box with a metal surface perforated by holes and straight or zigzagging lines in various sizes. The applicant was to hold a metal tack that he or she led in and out and along the hole shapes. Since the box and the metal tack were hooked up to an electrical device, a clock would ring out if the tack made contact with the metal sheet in the box, thus testing the applicant's dexterity and control.<sup>44</sup> The test, Eng wrote, should be carried out in a 'calm setting', since a lot of the applicants were nervous, and would be further rattled by the sound of the clock when they failed (a silent contraption would be preferable, she wrote, but thus far this had proved too costly). This test therefore came with significant statistical deviations and was wholly dependent on the subject's performance at this particular moment.

What truly separated goldsmithing from ordinary metalwork, Eng wrote, was that goldsmithing was an artisan craft, where artistic sensibilities played a much larger role. Here we can see how Eng's aforementioned interest in art and artistic expression factored into her psychotechnical work. To accommodate for this aspect, she therefore included a practical test where the applicants were to make a piece of metal jewelry, which was measured against a drawn figure, depicting the correct shape. They were then tasked with drawing a shape from memory, testing their 'sense of shape' as well as their memory. Another test of their sense of shape was conducted by a 'sorting test', also developed by Eng. Here, the participants had to sort different small metal pieces as quickly as possible. Then followed a test of their sense of scale, a test of their sense of weight, and a test of their perception of depth, all conducted with small metal pieces. Eng apparently strived to include the material of the work they had applied for in her test. Lastly, they had to judge samples of "good" and "bad" goldsmithing, thereby testing their aesthetic taste.<sup>45</sup> Typically for a psychotechnical test, various types of knowledge and judgments, ranging from psychomotorical measurements to discussions of aesthetic taste, intermingled in both the construction and the execution of the test.

---

44 Helga Eng, *Psykoteknisk prøve for elever ved Oslo guldsmedforskoole* (1926), pp. 1–2, [https://urn.nb.no/URN:NBN:no-nb\\_digibok\\_2017112148091](https://urn.nb.no/URN:NBN:no-nb_digibok_2017112148091)

45 Eng, *Psykoteknisk prøve for elever ved Oslo guldsmedforskoole*, pp. 3–4.



## Constructing the Test: Practical Knowledge and Psychological Theory

What role did intelligence or mental age play in this context? Both Eng's written comments, as well as the numerical value apparently ascribed to the intelligence test, indicate that it played a relatively subordinate role in the test battery for the goldsmith test. One can wonder why it was there at all. Without clear evidence, all we can do is speculate. The test seemingly had two potential functions. The first was to ensure that the applicants' mental faculties were "normal" and not significantly below their age group. While no one in the experiment Eng describes scored well below average, one could speculate that a very low score would lead to a disqualification from the application process altogether. The second function has to do with scientific legitimacy. As most of the subtests were practical, we can speculate that Eng wanted to include a type of test which was wholly "psychological", such as the intelligence test. This would ensure that the testing remained within the jurisdiction of the psychologist, in addition to providing the school with the types of tools—psychotechnical and scientific—they requested. As the legitimacy and areas of application of intelligence tests were growing, this test could be a stand-in for scientific psychology as a whole. Both of these explanations are plausible, and not necessarily contradictory. More generally, the inclusion of the intelligence test in this context shows how these types of tests became integrated into an ongoing psychological practice.

Eng's written account of how she constructed the test offers interesting insights into how the interplay between different types of knowledge, concerns, limitations, and expertise came to shape her psychotechnical practice.<sup>46</sup> In preparation for the construction of the test, she would visit the goldsmiths' workshop to better understand the processes involved. In doing this, she followed the advice of many theoreticians of psychotechnics who stressed that a thorough study of the

---

46 This, however, comes with an important caveat: although she often refers to the practical and professional knowledge and discretion of the teachers at the school, only her voice is directly discernible in the text. It is therefore possible that she only paid lip service to these teachers, whilst constructing the test solely based on her own discretion. Although possible, I find this to be unlikely, and will not discuss it further. It is however important to keep in mind.

work was needed to develop tests. Eng also conferred with the teachers, interviewing them about goldsmithing and gaining insight into what they deemed to be necessary skills. When the testing had begun, the test results of the successful applicants were checked against the teacher's grades after a few months at the school. Eng was pleased to find a high degree of correlation, and, according to her, the teachers were also happy with the results. The test for goldsmiths would become a mainstay at the Vocational School and was often the test Eng emphasized in her public lectures and texts on psychotechnics.

The codification, or translation, of manual and artistic skill into a psychologically testable set of attributes was thus undertaken in apparent collaboration with non-psychologists. This illustrates how psychotechnical tests, and perhaps psychological tests in general, were shaped by an interplay of pragmatic and scientific concerns. The specificity of the goldsmith tests meant, however, that its results were difficult to transfer or transpose to other contexts. The intelligence test, and its IQ score, have historically shown a greater degree of flexibility in this regard, where the results could be interpreted and reinterpreted in different societal, institutional, and scientific contexts. While the design and the mechanics of the tests are (superficially) similar, the character of the "object" measured in an intelligence test and a psychotechnical test such as Eng's is arguably different. They are both made up of an array of subtests producing scores that are added up to one final score; however, the "mental age", "intelligence" or "intelligence quotient" produced in an intelligence test contains an epistemic claim to universality. "Intelligence", for instance, was widely understood to be a fundamental trait determining a whole set of attributes, while the score from the psychotechnical test only indicated the specific proclivity for the type of work being tested for. Whether one accepted the fundamental existence of "intelligence", or "g", the number was still applicable in several settings.<sup>47</sup>

Thus, intelligence, at least in many contexts, was an object that could travel larger distances, and be used and reused due to its apparent

---

47 Although the existence of "intelligence" as a real, discernible object was contested already in the 1920s, I still find it appropriate to argue that it nonetheless had a more "real" and thus universal quality than the score from a psychotechnical test. An IQ score, whether or not it was taken to reflect upon a real thing, could still be interpreted and reinterpreted in many different situations and institutional settings.

fundamentality. Although both an intelligence test and a psychotechnical test, such as the one for goldsmiths, create a hierarchy within a particular group, they differ both in scale and in the posited universality of this hierarchy. Furthermore, the objects they measure are different with regards to the institutional setting—while intelligence, or mental age, mainly reflect scholarly aptitudes in accordance with the demands of ordinary schooling, the psychotechnical test measures cognitive and manual abilities against the specific demands of a certain type of work.

### Part of the Municipal Labor Program for Youths and Circulation in Popular Press

The development in the 1930s shows how psychotechnics became more well-known in Norway, as well as how it was grafted onto existing institutional practices and discourses. This entailed a growing societal role and legitimacy for psychologists. In 1932, the Psychotechnical Institute was incorporated into the Oslo municipality's labor program for youths. In an attempt to address the growing unemployment among people in the age range from fifteen to approximately twenty-four, the labor program emphasized vocational counseling. This counseling was, as far as the sources indicate, not performed by trained psychologists.<sup>48</sup> However, the program itself was, from 1935, headed by Dag Bryn, who had received his *magistergrad* [MPhil degree] in psychology in 1932 and had been employed as an assistant at the Psychotechnical Institute during the late 1920s and early 1930s. It is perhaps due to this connection that the institute became a part of the program. Specifically, an applicant in the labor program might, after vocational counseling, be sent to the institute for testing if they had been deemed suitable for skilled work in an industry or handicraft. Thus, the institute became incorporated into a larger system of counseling young workers.<sup>49</sup>

48 Helga Eng, *Norsk psykoteknikk i ti år 1925–1935: historisk oversikt* (1935), p. 13 [https://urn.nb.no/URN:NBN:no-nb\\_digibok\\_2014112738005](https://urn.nb.no/URN:NBN:no-nb_digibok_2014112738005); Lønnå, *Helga Eng: Psykolog Og Pedagog i Barnets Århundre*, pp. 160–61.

49 From the early 1900s, psychotechnics had an intertwined, although sometimes complicated relationship with the vocational guidance movement. In the 1910s, Hugo Münsterberg collaborated for a short while with the “father” of the US career guidance movement, Frank Parsons. The collaboration, however, ended in disappointment. In the 1920s, the “hard science” of psychotechnics was sometimes

With regards to professionalization, this can be regarded as another important step. While psychology, through Eng and her assistants' practice of psychotechnics, was used in a specific institutional context at the Vocational School, the integration into the municipal labor program in Oslo integrated this practice and its practitioners, the psychologists, into a larger public apparatus. Whether psychotechnics was regarded as indispensable to this program is hard to determine. But this integration did grant psychotechnics an even greater legitimacy, in addition to strengthening the connection between youths, education, psychology, and psychotechnics in Norway. Compared to the lofty ideas of the 1920s, exemplified by the program proposed by Keilhau, Hansen, and Schjelderup in 1921, psychotechnics came to have a more specific and limited meaning during the 1930s.

In the 1930s, psychotechnics received increasing attention in Norwegian publications. It became increasingly associated with problems associated with youths, puberty, and "choosing a path in life". In various popular books<sup>50</sup> that contained practical and sometimes philosophical guidance on how to choose a career path, psychotechnics gradually emerged as one of the techniques available both to young people and their parents in choosing and working towards a future vocation.<sup>51</sup> This growing interest, and the integration of psychotechnics into these types of books and pamphlets indicate an increasing cultural acceptance and legitimization of psychotechnical intervention, and the general ethos surrounding the application of scientific methods for governing individuals. Moreover, the task of guiding and measuring aptitude for work, and the task of sorting people into career paths, increasingly came to be associated with the psychological profession, thus granting it a clearer identity and jurisdiction.

---

differentiated from the "soft science" of guidance. In Norway, the relationship was intertwined, and the two sometimes confused. Helga Eng, for instance, held the belief that psychotechnics was best put to use as a specialized part of general counseling.

50 This genre of books seems to have developed during the early 1900s and became increasingly widespread during the 20s and especially 30s.

51 Alf Krohn, *Hvad Skal Jeg Bli? De Økonomiske Muligheder Inden Forskjellige Erhverv* (J. W. Cappelens Forlag, 1915); 'Norsk forening for socialt arbejde', *Hvad Skal Våre Døtre Bli?* (Olaf Norlis Forlag, 1928); K. F. Dahl, 'Yrkesveiledning Med Psykotekniske Prøver', *Socialt Arbeid*, 6 (1932), 81–92; Helga Eng, 'Yrkesveiledning', in *Foredrag Ved 8de Landsmøte i Oslo 22–26. September 1929* (Webergs boktrykkeri, 1930), pp. 65–76; Georg Brochmann, *Jeg Vil Bli Til Noe* (J. W. Cappelens Forlag, 1938).

This jurisdiction arguably came at the expense of parents. As with the alleged erosion of the relationship between master and journeyman in handicrafts and artisanal work that was put forward as a rationale for the implementation of psychotechnics at the Vocational School, parents' capability of providing solid advice and guidance to their children with regards to employment increasingly came under scrutiny during the 1930s.<sup>52</sup> This was not only due to some parents' susceptibility to an overestimation of their children's capabilities, but also due to rapid technological and demographical changes. Whereas work was often inherited from parents in earlier generations, this became uncertain due to some jobs vanishing, while others necessitated new forms of expertise and a greater degree of specialized skills. Nevertheless, psychotechnics here were mostly employed in a superficial sense, albeit often with explicit reference to the work carried out at the Psychotechnical Institute. There are few explicitly scientific, methodological, or theoretical discussions on psychotechnics in these books. The malleability of what was meant by "psychotechnics" thus persisted, and arguably helped the spread of psychotechnical knowledge, given that the diverse cast of stakeholders could still fill it with their own content. Still, the association between psychologists as a group and as a profession, and the tools of the psychotechnical test, arguably grew stronger due to this popularization—in part helped by Eng herself, who published extensively in various popular journals and newspapers.<sup>53</sup> Thus, we can say that psychotechnics became embedded in a larger institutional and discursive framework during the 1930s, which perhaps made the practices more resilient.

### Henry Havin: Intelligence and Psychotechnics at Odds in the Late 1930s and 1940s

Before I conclude, I will touch briefly on some key developments at the Psychotechnical Institute during the 1930s. Although the IQ test and psychotechnics had shared a more or less peaceful coexistence during

---

52 Dahl, 'Yrkesveiledning Med Psykotekniske Prøver', p. 82; Anathon Aall et al, *Pedagogisk Psykologi, Arbeidsmåten i Folkeskolen: Håndbok for Lærere* (Gyldendal, 1933), p. 153; Eng, 'Yrkesveiledning', p. 66.

53 Eng, 'Yrkesveiledning'; Eng, *Norsk psykoteknikk i ti år 1925–1935*.

much of the 1920s and 1930s, some fundamental rifts started to appear in the Norwegian scene during the late 1930s and 1940s. They now came to be thought of as a part of the same type of measurement and practice, sharing the goal of rationalizing society and making it more productive.

Eng left the Psychotechnical Institute to become Norway's first professor in pedagogy at the University of Oslo, and her long-time assistant Henry Havin took over the leadership of the institute. Havin had attained his MPhil in psychology in 1929, having specialized in the study of religion. His inaugural lecture, however, was on personality and personality typologies—a type of psychology he would continue to develop and make use of throughout his career.<sup>54</sup> In his psychological work, which he conducted at the institute as well as in his private time, he became increasingly interested in utilizing personality typologies in psychotechnical testing.<sup>55</sup>

Eng's departure and Havin's instatement marks a shift in the history of Norwegian psychotechnics, which came with greater ambitions on behalf of both psychotechnics as a science, and psychology as a profession. Havin showed skepticism towards intelligence testing, insofar as it was used in isolation, and not together with other tests. In 1938, Havin published an essay in *Norsk pedagogisk tidsskrift* [the Norwegian Pedagogical Journal], entitled 'Better than intelligence tests', where he criticized these tests—among them the Binet–Simon in particular—for belonging to an 'intellectualistic' phase of psychology best left behind. Rather than merely testing intelligence, he believed that psychotechnical tests should measure theoretical and practical aptitudes. Expressing frustration over the apparently widespread notion that psychotechnics only measured 'simple' abilities, such as fine motor skills, Havin pointed out that the inclusion of 'characterological' measurements would prove more valuable in predicting future success in employment than intelligence testing alone. With criticism apparently directed towards

54 Henry Halvorsen, 'Personlighetstyper', *Vor Verden*, 6.5 (1929), pp. 261–72.

55 Halvorsen, 'Personlighetstyper'; Henry Havin, *En psykoteknisk undersøkelse av sykepleierske-elever*, *Norbok*, Publikasjon (trykt utg.) [1] (O. Fredr. Arnesen, 1941) [https://urn.nb.no/URN:NBN:no-nb\\_digibok\\_2014102438024](https://urn.nb.no/URN:NBN:no-nb_digibok_2014102438024); Henry Havin, *Fortsatte psykotekniske undersøkelser av sykepleierske-elever*, *Norbok*, Publikasjon (trykt utg.) 3 (Oslo arbeidskontors psykotekniske institutt, 1942) [https://urn.nb.no/URN:NBN:no-nb\\_digibok\\_2014040408337](https://urn.nb.no/URN:NBN:no-nb_digibok_2014040408337); Henry Havin, *Karakterisering: et karakterologisk grunnskjema med ordlister og hefte for (selv-)karakterisering*, *Norbok* (O.F. Arnesen, 1944) [https://urn.nb.no/URN:NBN:no-nb\\_digibok\\_2007062600028](https://urn.nb.no/URN:NBN:no-nb_digibok_2007062600028)

teachers, he claimed that correctly conducted psychological testing in general, and especially intelligence testing, was dependent on the expert discretion of the psychologist.<sup>56</sup>

Havin was also a proponent of expanding the practice of psychotechnics into ‘more complex’ jobs, and for sorting out applicants to higher education. His ambition was thus to broaden the scope of psychotechnics, and potentially supplant intelligence testing in certain contexts with a psychotechnical program. The article from 1938 is one of the first examples to show psychotechnics pitted against intelligence testing. From being closely associated in a general test regime, this indicates how they conceptually came to grow apart. The interest in personality among psychotechnicians both in Norway as well as in Europe might provide an explanation. The movement away from “instrumentalistic” and atomistic measurements towards a holistic evaluation of the whole personality might have led to a more open conflict with already established practices of intelligence testing by teachers, psychiatrists, and others. This example shows that psychologists such as Havin had become mindful of the loosely defined jurisdiction over these types of tests. Now defined as a tool that was best utilized in the hands of a trained psychologist—not alone but in conjunction with other (and sometimes better) tools—the intelligence test, and testing in general, was to become the basis from which this new profession gained a legally and socially recognized jurisdiction.

## Conclusions

As we have seen, psychotechnics granted psychologists a new type of employment outside the university system during the 1920s. The Psychotechnical Institute proved a training ground for a professional practice for psychologists. Two of Eng’s assistants, Havin and Bryn, were some of the first to receive a degree in psychology from the University of Oslo, and among the first to gain employment as psychologists (or due to their psychological training) outside academia. Psychotechnical practice thus strengthened the nascent professional identity of psychologists in Norway and gave them a degree of control over a field

---

56 Henry Havin, ‘Bedre Enn Intelligensprøver’, *Norsk Pedagogisk Tidsskrift*, 22 (1938), pp. 74–75.

of intervention, namely the testing of youths for different kinds of artisan and manual work. The spread of psychotechnical knowledge also led to a demarcation of the professional jurisdiction of the psychologist, laying the groundwork for a professionalization which, in many ways, culminated in the establishment of a professional exam in psychology at the University of Oslo in 1948.

Intelligence testing was a part of psychotechnical practice. It played a role in the test battery for goldsmiths developed by Eng, and in many other similar tests, both in Norway and in Europe. The role it played was arguably a subordinate or supplementary one. Still, its inclusion in these tests, and their continued use, further proliferated intelligence testing, and its role as a supplementary part of a testing practice contributed to its staying power. The growing legitimacy and spread of psychotechnics carried the intelligence test into new areas, where it became a part of an established and continued practice. Psychology and psychological testing are still a part of the labor market today, through occupational and organizational psychology, and they became a part of this field through psychotechnics.

Returning to the white paper that established the professional exam in psychology in 1948, we can see that the gradual but important development of psychotechnical testing, and the increasing association between testing and the psychologist, led to a legitimization and legal recognition. Whether the testing—discussed in the white paper by prime minister Gerhardsen at the psychological conference in 1947, and in the parliament—was the same type of testing is, however, unclear. It probably was not, at least not in the technical sense. In preparation for the parliamentary proposition, a text written by psychologist Vidkun Coucheron-Jarl was sent to the members of the Norwegian psychological association. Here, he stressed the importance of attaining conceptual clarity, and keeping control over the term “psychotechnics”, which, according to him, had become devoid of any special, technical meaning due to widespread misuse.<sup>57</sup>

This shows that psychotechnics had become widely known, although poorly defined, and that psychologists tried to strengthen their jurisdiction over it in preparation for the establishment of the professional

---

57 Vidkun Coucheron-Jarl, ‘Ms. Fol. Brev Nr. 652 “Om ‘psykoteknikk’ i Psykologisk Terminologi”’, 1946.



exam. While psychology, psychotechnics, and psychological testing remained relatively uncontroversial in Norway during the interwar years, at least when they concerned vocational training and the labor market, some controversies started to emerge after World War Two.<sup>58</sup> In the short but interesting public debate that followed the publishing of the white paper, skepticism was levelled against psychological testing due to the apparent power it granted the psychologist over individual fates. This criticism, which was put forward by theologian Hans Ording in *Morgenbladet*, was rooted in a skepticism towards intelligence testing, and the negative consequences he had observed in the case of a young girl who had received a (probably too) low IQ score at school, and thus had lost all faith in herself.<sup>59</sup> Schjelderup, among others, defended the proposition by stating that incorrect or unnuanced testing, such as the example Ording was referring to, was the exact thing that the professional exam would prevent. A stronger degree of control over the testing, as well as a strengthening of the role of the discretionary judgment of the psychologist, would stop the tests from being misused. Whether Schjelderup meant that they would be prevented from being misused by other professions, such as teachers, is not explicitly stated, but it is reasonable to assume that he, and other psychologists, felt that control over the tests belonged to them.<sup>60</sup>

Controversies aside, psychological testing was here to stay, and is still a bedrock psychological practice in Norway and elsewhere. Even though the term “psychotechnics” faded from use quickly after World War Two, its core tenets are still present in today’s organizational and industrial psychology. As we can see, both the legitimization of and the critique against psychology in the late 1940s were rooted in the idea of psychology as a testing practice that had little to do with clinical treatment. Although psychology would become a clinical profession in Norway, it was founded partly in psychotechnical testing in the labor market.

---

58 On the controversies surrounding intelligence testing in Norway after World War Two, see chapters 11 and 4, this volume.

59 *Morgenbladet*, 23.08.1947, 3

60 *Morgenbladet*, 27.08.1947, 2–3

## Bibliography

- Aall, Anathon, Harald Schjelderup, H. P. Barden, and B. Ribsskog, *Pedagogisk Psykologi, Arbeidsmåten i Folkeskolen: Håndbok for Lærere* (Oslo: Gyldendal, 1933).
- Baklien, Bergljot, *Psykologprofesjonen Og Dens Omgivelser, Utredninger Om Forskning Og Høyere Utdanning*, 6 (Oslo: NAFs utredningsinstitutt: Norges almenvitenskapelige forskningsråd, 1976).
- Blatter, Jeremy Todd, 'The Psychotechnics of Everyday Life: Hugo Münsterberg and the Politics of Applied Psychology, 1887–1917' (Doctoral dissertation, Harvard University, 2014), <https://dash.harvard.edu/handle/1/12274284>
- Brochmann, Gerog, *Jeg Vil Bli Til Noe* (Oslo: J. W. Cappelens Forlag, 1938).
- Coucheron-Jarl, Vidkunn, 'Ms. Fol. Brev Nr. 652 "Om 'psykoteknikk' i Psykologisk Terminologi"', 1946.
- Dahl, K. F., 'Yrkesveiledning Med Psykotekniske Prøver', *Socialt Arbeid*, 6 (1932), 81–92.
- Eng, Helga, *Abstrakte Begreper i Barnets Tanke Og Tale. Psykologiske Undersøkelser Paa Grundlag Av Iagttagelse Og Ekspe-Rimenter Med Skolebarn* (Kristiania: Aschehough, 1912).
- Eng, Helga, *Norsk psykoteknikk i ti år 1925–1935: historisk oversikt* (1935), [https://urn.nb.no/URN:NBN:no-nb\\_digibok\\_2014112738005](https://urn.nb.no/URN:NBN:no-nb_digibok_2014112738005)
- Eng, Helga, *Psykoteknisk prøve for elever ved Oslo gultsmedforskole*, 1926, [https://urn.nb.no/URN:NBN:no-nb\\_digibok\\_2017112148091](https://urn.nb.no/URN:NBN:no-nb_digibok_2017112148091)
- Eng, Helga, 'Yrkesveiledning', in *Foredrag Ved 8de Landsmøte i Oslo 22–26. September 1929* (Oslo: Webers boktrykkeri, 1930), pp. 65–76.
- Gundlach, Horst, 'Psychology as Science and as Discipline: The Case of Germany', *Physis; Rivista Internazionale Di Storia Della Scienza*, 2006, 61–89.
- Halvorsen, Henry, 'Personlighetstyper', *Vor Verden*, 6.5 (1929), 261–72.
- Havin, Henry, 'Bedre Enn Intelligensprøver', *Norsk Pedagogisk Tidsskrift*, 22 (1938), 74–75.
- , *En psykoteknisk undersøkelse av sykepleierske-elever*, *Norbok*, Publikasjon (trykt utg.) [1] (O. Fredr. Arnesen, 1941), [https://urn.nb.no/URN:NBN:no-nb\\_digibok\\_2014102438024](https://urn.nb.no/URN:NBN:no-nb_digibok_2014102438024)
- , *Fortsatte psykotekniske undersøkelser av sykepleierske-elever*, *Norbok*, Publikasjon (trykt utg.) 3 (Oslo arbeidskontors psykotekniske institutt, 1942), [https://urn.nb.no/URN:NBN:no-nb\\_digibok\\_2014040408337](https://urn.nb.no/URN:NBN:no-nb_digibok_2014040408337)
- , *Karakterisering: et karakterologisk grunnskjema med ordlister og hefte for (selv-)karakterisering*, *Norbok* (O.F. Arnesen, 1944), [https://urn.nb.no/URN:NBN:no-nb\\_digibok\\_2007062600028](https://urn.nb.no/URN:NBN:no-nb_digibok_2007062600028)

- Helsvig, Kim, *Pedagogikkens Grenser: Kampen Om Norsk Pedagogikk Ved Pedagogisk Forskningsinstitutt 1938–1980* (Oslo: Abstrakt Forlag, 2005).
- Høstmælingen, Andreas, 'Hvordan Norsk Psykologi Ble En Klinisk Profesjon', *Tidsskrift for Norsk Psykologforening*, 59.4 (2022), 248–67.
- Krohn, Alf, *Hvad Skal Jeg Bli? De Økonomiske Muligheter Inden Forskjellige Erhverv* (Kristiania: J. W. Cappelens Forlag, 1915).
- Lønnå, Elisabeth, *Helga Eng: Psykolog Og Pedagog i Barnets Århundre* (Fagbokforlaget, 2002).
- Meskill, David, *Optimizing the German Workforce: Labor Administration from Bismarck to the Economic Miracle* (New York: Berghahn Books, 2010).
- Nilsen, Håvard Friis, *Du må ikke sove: Wilhelm Reich og psykoanalysen i Norge*, 2. opplag (Oslo: Aschehoug, 2023).
- , 'Psykologene: Universitetsfaget Som Ble En Klinisk Profesjon', in *Profesjonshistorier* (Oslo: Pax, 2014), pp. 450–70.
- , *Resistance in Therapy and War: Psychoanalysis in Norway 1920–1945 (Doktorgradsavhandling)* (Oslo: Universitetet i Oslo, 2016).
- Norsk forening for sosialt arbeide, *Hvad Skal Våre Døtre Bli?* (Oslo: Olaf Norlis Forlag, 1928).
- Rabinbach, Anson, *The Eclipse of the Utopias of Labor* (New York: Fordham University Press, 2018).
- Triantafillou, Peter, and Alfonso Moreira, 'Modern Templates of Happiness: Performing Spiritualism and Psychotechnics in Denmark', *History of the Human Sciences*, 18.2 (2005), 87–109.

## 6. From Segregation to Integration and the Role of Testing in a Norwegian Educational Psychology Office, 1953–1980

*Kari Ludvigsen*

---

Educational psychology services have been regarded as a persistent stronghold for psychometric testing since the interwar period in a range of previous studies. This chapter deals with disruption and continuity in the approach towards the use of IQ tests and other assessment tools in the Norwegian Educational Psychological Counselling Services (EPS) [*Pedagogisk-psykologisk tjeneste, PPT*] from the early 1950s until 1980. The analysis is based on the scrutiny of yearly reports from the municipal EPS office in Norway's second largest city, Bergen. How was testing, individually and in groups, described and legitimized in the reports throughout this period? Which tensions regarding testing can be traced, and how can these be understood in relation to the educational reforms and changing role of the EPS services?

### Understanding the Assessment of Schoolchildren in EPS

A range of studies have assessed how the development of educational testing has influenced concepts of normality and deviance in childhood and upbringing. Psychometric testing within an educational context has been linked to the growth of what has been labelled “psy- expertise”, as well as to the development of a global testing culture and applied psychology.<sup>1</sup>

---

<sup>1</sup> Nikolas Rose, *Powers of Freedom. Reframing Political Thought* (Cambridge: Cambridge University Press, 1999); Karen E. Andreasen and Kristine Kousholt, ‘Minorities and educational testing in schools in Arctic regions: An analysis and discussion focusing on normality, democracy, and inclusion for the cases of Greenland and the Swedish Sami schools’, in *Testing and Inclusive*

Further, international research into the assessment of school children stresses that data from psychometric testing contributes to shaping discourses on education and influences ideas and aims of schooling and teaching.<sup>2</sup> The expansion of IQ testing from the interwar period has been regarded as crucial to disciplining and systematizing the population by producing distinctions between gifted, medium gifted and less gifted children.<sup>3</sup> Several studies stress the role of education psychology in differentiating and defining childhood normality and deviance, in a Scandinavian setting as well. According to Bjørn Hamre and Christian Ydesen, ideas from psychology, social engineering, eugenics and pedagogy merged into efforts of securing a better future by categorizing children during what they label 'the age of measurement', from the 1930s onwards. Danish educational psychology sought a hold of IQ testing to handle risks related to population degeneration and optimizing of the school system.<sup>4</sup> The study by Hamre and Ydesen points to a disruption in the role of IQ testing in education from the 1970s, as technologies of differentiation gradually became replaced by what they label security measures, aimed at optimizing population potential and reducing risks. They regard the evolving ideas and practices of inclusive education as part of national competitiveness through education. This provided a new role for education psychology to promote ideas of individual development and elevate students to become reflexive, flexible, knowledgeable, and self-reliant individuals.<sup>5</sup>

In the Norwegian case, Jan Froestad and Bodil Ravneberg argue that public policies and administrative and professional practice formed disability as a category in relation to distinct national conditions for modernization.<sup>6</sup> This and several other studies have stressed that integration of children from different social classes within a Norwegian unitary primary school involved

---

*Schooling: International Challenges and Opportunities*, ed. by Bjørn Hamre, Anne Morin and Christian Ydesen (London: Routledge, 2018), pp. 19–33; Christian Ydesen and Karen E. Andreasen, 'Historical Roots of the Global Testing Culture in Education', *Nordic Studies in Education*, 40 (2020), 149–166. <https://doi.org/10.23865/nse.v40.2229>.

2 Sverker Lindblad, Daniel Pettersson and Thomas S. Popkewitz, eds., *Education by the Numbers and the Making of Society. The Expertise of International Assessments* (New York: Routledge, 2018), <https://doi.org/10.4324/9781315100432>.

3 Stephen J. Ball, *Foucault, Power and Education* (Routledge 2013), p.68; Bjørn Hamre and Christian Ydesen, 'The Ascent of Educational Psychology in Denmark in the Interwar Years', *Nordic Journal of Educational History*, 1, no. 2 (2014), 87–111, <https://doi.org/10.36368/NJEDH.V1I2.40>.

4 Hamre & Ydesen, *The Ascent*, p. 106.

5 Hamre & Ydesen, *The Ascent*, pp. 107–108.

6 Jan Froestad and Bodil Ravneberg, 'Education Policy, the Norwegian Unitary School and the Social Construction of Disability,' *Scandinavian Journal of History*, vol. 31, issue 2 (2006), 119–143, <https://doi.org/10.1080/03468750600604119>.

segregation of other categories of children, and services and expertise aimed at classifying children not suitable for ordinary education has particularly been regarded as a decisive element to fulfil the 1950s education reform.<sup>7</sup>

The first Norwegian EPS offices were established in Oslo and Aker in the late 1940s. From the 1950s onwards, the services gradually expanded nationwide to serve the differentiated education system, with testing and advising of suitable educational measures as a main task. New approaches in special needs education promoted principles of inclusion and normalization from the late 1960s, reflected in new education acts in 1969 and 1976.<sup>8</sup> Parallel to these reforms, the role and expectations of the EPS services seem to have broadened and changed. Some historical analyses have pointed to tensions in the approaches and methods of the EPS, including when it comes to the role of testing.<sup>9</sup> According to Dalen, the segregation approach in special needs education gave way to ideas of integration from the mid-1960s; however, she stresses that this turn of tide must not be regarded as a fixture.<sup>10</sup> More recent studies have underlined the continuing crucial role of IQ testing in the EPS services—despite the shifting legitimacy of such assessment tools—changing education policies and subsequent redefinitions of the responsibilities, tasks, and expertise of counselling services.<sup>11</sup>

Thus, previous studies point to both persistence and change in the work of EPS offices in general, and in the role of IQ testing in particular, in the decades after World War Two. This makes it relevant to look closer at the presentations of psychometric testing in the EPS of one of Norway's largest municipalities in a period marked by changing approaches towards special needs education, from differentiation to inclusion. How is IQ testing presented and legitimized, and which continuities and disruptions can be traced?

7 Froestad and Ravneberg, *Education Policy*; Eva Simonsen, *Vitenskap og profesjonskamp. Opplæring av døve og døvsskole i Norge 1881–1963* (Unipub forlag 2000); Marianne Sandvik Tveitnes, *Sakkyndighet med mål og mening: En analyse av sakkyndighetens institusjonaliserte kjennetegn; et grunnlag for refleksjon og endring* (unpublished doctoral dissertation, University of Stavanger, Norway, 2018); Peder Haug, 'Ideal og realitet i spesialundervisninga', in *Pedagogikkens idé og oppdrag*, ed. by Harald Thuen, Synnøve Myklestad and Stine Vik (Bergen: Fagbokforlaget, 2022), Chapter 7.

8 Hans J. Gjessing, 'Norsk spesialundervisning gjennom tre decennier 1950–1980', *Spesialpedagogikk* 3 (1972), 2–13; Monica Dalen, *Integrering av funksjonshemmede i grunnskolen - en prinsipiell drøfting* (Oslo: Universitetsforlaget, 1985), pp. 27–36.

9 Geir Evjen, 'Pedagogisk- psykologisk rådgivning- et historisk perspektiv', *Monography, Skolepsykologi, Materiellservice* (1984); Oddvar Vormeland, *Pedagogisk-psykologisk rådgjeving i skolen* (Fabritius, 1963).

10 Dalen, *Integrering*, p.36.

11 Simonsen, *Vitenskap og profesjonskamp*; Tveitnes, *Sakkyndighet*, Haug, *Ideal og realitet*.

The chapter's analysis is based on annual reports to the municipal education authorities from the Bergen EPS office, normally authored and signed by the head of office.<sup>12</sup> The reports differ in length, form, and content, and stress various aspects of the work conditions, organization, tasks, and priorities. They give accounts of the service content, staffing and organization, the number of users, and some of the methods and tools involved in the work. The annual reports contain little detailed data on the assessment of individual school children, and except for the first few years, they do not provide a systematic account on the use of specific testing tools. These documents do, however, provide a picture of the changing priorities and their justifications towards the local authorities, and contain descriptions of preferred work methods, including individual and group testing. Kristin Asdal has suggested what has been labelled a practice-oriented document analysis, where documents are not merely regarded as descriptions of an external reality, but as forms of social action and contributing to shaping that reality.<sup>13</sup> This approach is highly relevant in studying reports written by heads of staff in public services. Such documents will often follow certain genre standards and are formally required by the authorities, like the annual EPS reports analysed here. The reports were produced by heads of office, who have a certain authority due to their expertise and position. They may also reflect certain agendas towards the local authorities who fund and oversee the services, and thus also contribute to producing certain standards for the services. These factors render these reports a relevant source for tracing changing approaches to the role of different testing tools and measures in the emerging EPS system, and their social justification.

---

12 In 1972, the office was reorganized in connection with municipal mergers. The offices in four former suburban municipalities were converted into district offices under joint management with the former Bergen EPS office. The following years, the annual reports were written by the district psychologists. The analysis here is based on the annual reports from the Bergen city center district.

13 Kristin Asdal and Ingunn Moser, 'Experiments in context and contexting', *Science, Technology & Human Values*, 37 (4), 291–306, <https://doi.org/10.1177/0162243912449749>; Kristin Asdal, 'What is the issue? The transformative capacity of documents', *Distinktion: Scandinavian Journal of Social Theory*, 16 (1), (2015), 74–90, <https://doi.org/10.1080/1600910X.2015.1022194>; Hilde Reinertsen and Kristin Asdal 'Calculating the blue economy: producing trust in numbers with business tools and reflexive objectivity', *Journal of Cultural Economy*, Vol. 12, issue 6 (2019), 552–570, <https://doi.org/10.1080/17530350.2019.1639066>.

## Norwegian Education Reforms and the Development of Education Psychology Services

The Norwegian EPS developed as part of the education system from the early 1950s, and its role and tasks have changed in line with reforms in the education sector. When it comes to testing, previous research has regarded such developments in relation to the transformation of special needs education—from a differentiated system of state special schools with associated specialized teacher competence during the 1950s, towards reforms aimed at integration of all children into the comprehensive educational system in the mid-1970s.

A state special education system was established in Norway through the 1881 Abnormal School Act [*Lov om abnorme Børns Undervisning*]. This has been regarded as an important vehicle for establishing comprehensive schools for children from all social classes.<sup>14</sup> Compulsory education for deaf, blind and “mentally retarded” children was organized into a differentiated system of separate state-funded boarding schools for groups of children considered unsuitable for education in ordinary classrooms.<sup>15</sup> From the early 1900s, city schools became further differentiated through the establishment of separate special education classes and schools for children considered weakly gifted, achieved through an alliance between municipalities, physicians, and progressive teachers.<sup>16</sup>

To assess children for special education measures, the municipalities of Oslo and Bergen engaged psychiatrists to carry out IQ testing from the 1920s. Additionally, a few mental hygiene clinics were established by key entrepreneurs within the field of psychology around Oslo during the 1930s, with an emphasis on assessment of the child’s family and surroundings, in addition to psychometric testing.<sup>17</sup> After World War Two, the largest cities established so-called counselling offices [*rådgivningskontor*], with the offices in Oslo and Bergen being some of the earliest.

14 Froestad and Ravneberg, Education Policy.

15 See Lefkaditou, chapter 1, this volume.

16 Bjørn Hamre, Thom Axelsson and Kari Ludvigsen, ‘Psychiatry in the sorting of schoolchildren in Scandinavia 1920–1950: IQ testing, child guidance clinics, and hospitalization’, *Paedagogica Historica*, 55(3) (2019), 391–415. <https://doi.org/10.1080/00309230.2019.1597137>; Froestad & Ravneberg, Education Policy.

17 Ludvigsen, Kari and Åsmund Arup Seip, ‘The establishing of Norwegian child psychiatry: ideas, pioneers and institutions. Arcadia’, *History of Psychiatry*, Vol. 20, No. 77 (2009), <https://doi.org/10.1177/0957154X08089847>; Christian Ydesen, Kari Ludvigsen and Christian Lundahl, ‘Creating an Educational Testing Profession in Norway, Sweden and Denmark, 1910–1960’, *European Educational Research Journal*, 12(1) (2013), 120–138, <https://doi.org/10.2304/eeerj.2013.12.1.120>.





Fig. 6.1 Åse Gruda Skard (1905–85) was a Norwegian psychologist who pioneered mental hygiene for schoolchildren, teachers, and parents from the 1930s. From her position in psychology at the University of Oslo, she was active in research and in international associations as well as a popular educator in childhood and upbringing. Photographer Bachrach 1941/National Library of Norway, PDM 1.0 (Public domain).

During the 1950s, the special education system was reorganized and professionalized, parallel to reforms in the ordinary education system. Prior to the 1959 elementary school reform, a new Special Needs Education Act in 1951 [*Lov om spesialskoler*] made way for growth in special needs education. A new state Directorate for the Special Schools [*Direktoratet for spesialskolene*] was empowered to make authoritative definitions of the educational needs for disabled children and extended differentiation of students. Through a 1955 amendment to the Elementary School Act [*Folkeskoleloven*], municipal

provision of special education became mandatory. The municipal special schools and classes were aimed at children who did not have severe enough difficulties to attend a special school but were considered unfit for regular classrooms. According to Froestad and Ravneberg, children defined as slow learners now became the major target for segregation.<sup>18</sup> The following years, both the state special school system and municipal special schools and classes expanded rapidly. In the school year of 1965–66, the number of special schools peaked at 58, serving a total of 2,500 students. The municipal special schools and classes grew from serving a few hundred children in the early 1950s to over 20,000 in the 1960s.

The main purpose of the EPS offices was to provide a more systematic foundation for selecting students for special classes and schools.<sup>19</sup> Following the new regulation of special needs education, the 1955 act gave provision to state co-funding for municipal school psychologist posts, aimed to carry out an 'examination of school maturity, aptitude, behavioural difficulties and academic difficulties of a more serious nature.'<sup>20</sup> The services were to provide advice on the transfer of schoolchildren to state special schools, municipal special education, or childcare and child psychiatry treatment. This new funding gave impetus to further development of counselling services during the following decades.<sup>21</sup> The services grew from 10 offices in 1956 to 103 in 1970, and 245 in 1981.

During the 1950s and 1960s, the guidelines for the EPS services were broad and the geographical variation in terms service organization and access to expertise was considerable.<sup>22</sup> In 1964, the newly established national Special Schools Council [*Spesialskolerådet*] proposed further development of the differentiated special school system. Psychometric testing was here considered a useful tool, not least regarding accurate categorization of mental retardation [*evnesvakhet*]. An incipient skepticism regarding the

---

18 Froestad & Ravneberg, Education Policy.

19 'Samordningsnemnda for skoleverket, oppnevnt ved kongelig resolusjon 7. mars 1947', *Sammenfatning og utsyn* (Mysen, Indre smaalenenes Trykkeri, 1952). The Coordination Committee for the education system [*Samordningsnemnda for skoleverket*] proposed that a pedagogical advisory service be established in each county. This was supported by the Department of Church and Education [*Kyrkje- og undervisningsdepartementet*] St.meld. nr. 9 (1954) *Om tiltak til styrking av skoleverket*).

20 Ot. Prp. Nr. 4, 1955. Om brigde i lov om folkeskolen på landet, lov om folkeskolen i kjøpstedene og lov om framhaldsskular.

21 NOU 1983:4, Pedagogisk- psykologisk rådgivningstjeneste (Flekkøyutvalget), 15–16.

22 NOU 1983:4, Pedagogisk- psykologisk rådgivningstjeneste (Flekkøyutvalget), 15.

validity of the established aptitude tests can, however, be traced in the report. It was advised that assessment should be based on investigations of the child's social abilities and the overall situation, in addition to different types of standardized tests. A few years later, the critical approach towards special needs education and testing practices in public and political debates and disputes amongst experts paved the way for education reforms, as well as experiments with new models for EPS work from the late 1960s.<sup>23</sup> Parallel to this, new models for EPS were tried out in several municipalities from the late 1960s, stressing preventive and local community work.

The 1969 Elementary School Act [*Grunnskoleloven*] included a new paragraph on counseling services, and introduced a new name for the service: *Pedagogisk-psykologisk tjeneste (PPT)* [pedagogical-psychological services].<sup>24</sup> The 1976 Primary School Act formed a new framework for the services.<sup>25</sup> All children were provided the right to educational services within the framework of the elementary school [*grunnskolen*] based on their abilities and prerequisites. An expert apparatus was regarded as a prerequisite for implementing the new act. The regulations specified the tasks for the services, and the EPS was now designated the role of an expert body for the kindergartens and schools in question regarding special needs. The scope of the services was broadened through inclusion of high school students [*videregående skole*] and pre-school children. The new act, however, did not specify the role of the EPS, and further development of the services depended upon county plans and a government white paper on special needs education in 1976. Although the lack of qualified personnel was reduced from the mid-1970s, staffing still showed large geographical variations.<sup>26</sup> This analysis embraces the period up until the early 1980s, when new regulations for the services were developed.

### Testing at the Forefront: Establishing the Bergen EPS Office

The school psychology office in Bergen was established in 1953, as one of the first in Norway outside Oslo. A full-time school psychologist and an office assistant were employed. The first psychologist, Håkon Østby,

23 See Thue, chapter 11, this volume.

24 Lov av 13. juni 1969 om grunnskolen, § 8.4.

25 Lov nr. 42 om brigde i lov av 13. juni 1969 nr. 24 om grunnskolen, datert 13. juni 1975.

26 NOU 1983:4, pp.15–16.

resigned in 1957, expressing disappointment that the state grant for school psychologist positions did not lead to expansion of the office. For the next decade, staffing and competence building were recurring issues nationwide, with limited access to expertise in psychology, social work, and special pedagogy. According to one report from the late 1950s, the office in Bergen lagged far behind all other municipal offices in the country in such aspects.<sup>27</sup>

The earliest reports describe the number of referrals and categorize cases. The office registered causes for referrals in two main categories: educational difficulties and behavioural challenges. In the first year, the distribution of cases was reported as rather even between these groups.<sup>28</sup> These categories do not, however, always seem to have been obvious to the staff, as the reports stressed an overlapping of the two main categories several times during the following decades.

The referrals included about twice as many boys as girls, and services aimed at the youngest schoolchildren continued as a main priority during the first decades. Teachers made most of the referrals, but the number of parents who initiated contact with the school psychologist increased. According to Østby's report, 162 children were examined during the first year. Out of these, 83 were examined individually and 79 by group test for school preparedness, and 21 cases were on the waiting list. For the next few years, the number of children referred was stable.

Pertaining to the specific use of psychometric testing tools, the annual reports stress this to a varying degree. In the first annual reports, the school psychologist Østby found it crucial to describe his efforts to get access to available testing tools, as well as their strengths and weaknesses. He categorized the tools into three groups: intelligence tests, personality tests, and educational tests.<sup>29</sup> Østby acquired the Norwegian edition of the Terman–Merrill test, as well as the Arthur Performance Scale I.<sup>30</sup> Amongst the personality tests, the Rorschach was

27 Annual report 1959–60, Skolepsykologisk kontor i Bergen, Bergen Byarkiv, Årsmeldinger, pedagogisk- psykologisk tjeneste. Sentrum. Arkivnr. A-3253, XA.

28 Annual report 1953–54.

29 Annual reports 1954–55, 1955–56.

30 In 1953, Østby reported that the Terman–Merrill test's Norwegian edition was delayed. As compensation, the office was allowed to take over school board member A. Barsnes' private copy of the Oslo test, which was no longer commercially available. He also collaborated with the headmaster at Årstad municipal remedy school on the preparation of the Hagerty–Olson–Wickman observation form, to make observations more uniform and easier. In 1955, the Oslo test had become fully replaced by Terman–Merrill (Annual report 1955–56).

regarded particularly useful, because it also provided a broad understanding of intelligence.<sup>31</sup> According to Østby, several tests were too complicated to use efficiently. He, however, reported positive experiences with some tests, in particular the Jackson's Family Attitude Scale, wherein the observation of children's play was facilitated by equipping an observation room with a Buhler Welt-Test said to provide possibilities for more systematic observations of play.<sup>32</sup>

The three first annual reports from Østby provided tables showing the use of the different testing tools at the office. Table 1 sums up the numbers in the period 1954–1957.

Table 1. Individual tests carried out from 1954–1957.

| Testing tools                     | Number of tests carried out |                                       |         |
|-----------------------------------|-----------------------------|---------------------------------------|---------|
|                                   | 1954–55                     | 1955–56                               | 1956–57 |
| The Oslo test <sup>33</sup>       | 18                          |                                       |         |
| Terman–Merrill <sup>34</sup>      | 57                          | 61                                    | 56      |
| Arthur Point Scale of Performance | 41                          | 46                                    | 45      |
| Other IQ tests <sup>35</sup>      | 45                          | 93<br>(including school preparedness) | 83      |
| Rorschach                         | 84                          | 95                                    | 77      |
| Other personality tests           | 24                          | 13                                    | 8       |
| Pedagogical tests <sup>36</sup>   | 3                           | 13                                    | 20 (+1) |
| TOTAL                             | 272                         | 321                                   | 290     |

These first reports show that individual tests were usually combined with a range of other tools, including observations of children playing and drawing, as well as interviews with children, parents, and teachers.<sup>37</sup> In the event of considering transfers to special needs schools,

31 Annual report 1954–55.

32 Ibid.

33 See Lefkaditou, chapter 1, this volume for a description of this test.

34 See Paulsen, chapter 9, this volume for a closer description of this test and its use.

35 The categories are based on the reports. The reports do not provide information on the content of the group labelled 'Other IQ tests'.

36 The reports list the category pedagogical tests, and describe the use of such tools, but with little detail. Specifically, a reading test named Bakke's test was mentioned.

37 Annual report 1954–55.

combinations of several individual tests were used. The most difficult cases were forwarded to the office of the municipal school psychiatrist, who decided on the right measures.<sup>38</sup>

Requests for testing the school preparedness of preschool children was an important force in establishing municipal and intermunicipal EPS offices in the 1950s,<sup>39</sup> and school preparedness assessment [*skolemodenhetsprøve*] was also a central topic in the Bergen EPS office reports from the 1950s and 1960s. Norwegian children were expected to start school in August the year they turned seven. The 1936 school laws, however, allowed minor children to attend school by application from parents to the school board.<sup>40</sup> In addition, the law (§ 13.3) allowed a child who turned seven years old during the calendar year to apply to postpone enrolment until the next term, i.e. after the eighth birthday.<sup>41</sup> This system was aimed at preventing children from experiencing learning challenges due to immaturity.<sup>42</sup>

Previous research has suggested that school preparedness tests may serve as a tool for identifying and promoting talent.<sup>43</sup> In this material, this kind of assessment seems to be regarded mainly as a tool for keeping immature children from starting school too early, as well as providing teachers with information on the maturity level of their students. School preparedness testing was tried out in the Bergen EPS from the start, and expanded into a more systematic practice during the early 1960s. This form of assessment was a central topic in reports until the mid-1970s, when the role of maturity testing diminished due to the integration of the Special Schools Act into the Primary School Act.<sup>44</sup> Sandven's maturity test from 1950 was the most common amongst a variety of

38 Annual reports 1953–54 and 1954–55.

39 Egil Gabrielsen and Kjersti Lundetræ, 'Indikerer de norske PIRLS- resultatene et behov for å justere retningslinjene for skolestartsalder?', in *Klar framgang. Leseferdighet på 4. og 5. trinn i et femtenårsperspektiv*, ed. by Egil Gabrielsen (Oslo: Universitetsforlaget, 2017), pp. 204–221, <https://doi.org/10.18261/9788215030258-2017-12>.

40 Children who turned six years old before the start of the school year in August.

41 Innst. O. XXIX (1936).

42 The Elementary School Act of 1969 adjusted the rules slightly. (Inst. O. XIV (1968–6)).

43 Axelsson, 'Rätt elev i rätt klass: skola, begåvning och styrning 1910–1950,' (Doctoral Dissertation, Linköping Studies in Arts and Science No. 379 Linköpings universitet, Institutionen för Tema Linköping, 2007); See Porter, chapter 13, this volume.

44 Gabrielsen and Lundetræ, *Indikerer*.

tests, used together with observations and conversations with parents.<sup>45</sup> This is also the only such test explicitly referred to in the Bergen EPS reports. The procedures of testing were described briefly in the first reports: a team comprising a psychologist and a trained teacher carried out the testing in groups of eight to fifteen preschool children over a two-day period, including observation of play during recess.<sup>46</sup> Parents filled out a form with questions about their children's development. At the initiative of parents or the psychologist, some preschool children were tested individually by use of the Oslo test and the Arthur scale. The activity was scaled up, and in 1955–56, group examinations of 307 pupils in 11 classes were carried out at two schools.<sup>47</sup>

With a new head of office from 1957, the content of reports from the Bergen EPS changed. In 1959, the new school psychologist Førde stated that the office was satisfactorily equipped with test material and literature, but that it still had to be kept up-to-date. Førde did not mention the number of tests carried out, nor did he list systematically the available testing tools like his predecessor had done. His reports briefly described the number of cases, and the procedures for handling them. Out of the eighty-three children reported to the office in 1959, sixty-three were invited to the office for individual examination one to four times each. According to the reports, most of the cases were discussed with the head teacher or teacher, and all parents were interviewed. The report also briefly mentions collaboration with representatives from the healthcare and childcare systems.

The 1959 report expanded more on the group tests. Arguing that standardized tests were objective aids that could provide hypotheses and further investigations, Førde described his initiative to use a group test battery, combining testing of school preparedness and proficiency in silent reading, dictation, and arithmetic.<sup>48</sup> The EPS distributed

---

45 See Helsvig, chapter 12, this volume regarding Sandven's work.

46 The school psychologist noted that some teachers regarded it as too much work to assist in the testing, and the job was also unpaid. The advantage of the experience was teachers gaining insights on the preparedness level amongst the first-graders, making it possible to adjust the teaching accordingly (Annual report 1956–57).

47 Annual report 1955–56.

48 The office administered standardized tests in Norwegian and arithmetic for all children from first and second grades in Bergen, based on two tests standardized in a Norwegian setting: Gjessing's Norwegian adaptation of Gates' diagnostic reading test, and an arithmetic test developed by J. Tornes. Children were further examined

the tests, trained participating teachers, and advised parents and the school.<sup>49</sup> However, there was no capacity for carrying out systematic group tests on a larger scale, and based on the results, Førde pointed out that there should be stricter requirements for the admission of minors to start school early. He regarded pedagogical difficulties as the main task of the service, and stressed the need for staff with specialist training and practice in clinical child psychology to handle the examination and effective treatment of children with severe adjustment difficulties.

The EPS reports from the 1950s focused on descriptions of individual assessment through a range of combined available testing tools, along with gradually expanding the testing of school preparedness and proficiency. They clearly reflect the tasks stressed in the guidelines and describe the establishment of individual as well as school preparedness and proficiency assessment. The first reports expanded on descriptions of the acquired testing tools, showing how the school psychologist built up a reservoir of up-to-date tools for assessment—both IQ tests and personality tests. While the first reports discussed the usefulness of the various tests and provided statistics on their use, the reports from the late 1950s provide far less detail on these topics and put more weight on collaborating with teachers on group tests. From the start, the EPS heads were concerned about the need to use a range of testing tools, and combining individual testing with observations and parent interviews to advise the right placement of the child. Furthermore, a tension appears between the categorization of children's problems into either learning difficulties or behavioral problems, in accordance with the differentiated special education system.

## Towards a More System-Oriented Approach

During the 1960s, the content of the EPS reports changed, with increasingly more weight on collaborative work with teachers, and efforts to restrict the scope of individual assessment. This may, in part, reflect the expanding workload and vacancies amongst staff, but also a more critical approach towards the role of testing tools.

---

for two days in groups of eight to ten, with the use of Preparedness Test I from PFI (Sandven's test).

49 Annual report 1959–60.



In 1962, a reorganization and expansion of the Bergen EPS was decided by the municipality, with positions for four school psychologists, two social counsellors, and two office assistants. In addition, the office was assigned access to sixteen hours of municipal child psychiatric expertise.<sup>50</sup> Under the lead of Gunnar Rakvåg, the office expanded and changed during the next decade.

Contrary to his predecessors, Rakvåg's reports did not provide detailed information on the use of individual tests. However, the number of individual students examined was reported, together with overviews of advisory measures. Descriptions of group testing and collaborative efforts were at the forefront. From the mid-1960s, the office prioritized systematic work at schools with the youngest schoolchildren, with hopes of a preventive effect. The reports also stressed testing of school preparedness, as well as proficiency testing in reading and math.

The resources for the municipal school psychology service expanded during the 1960s, but the increasing workload, financial restrictions, and lack of qualified personnel were recurring themes in the reports throughout the decade. According to Rakvåg, pressure from a growing waiting list made it difficult to stress the therapeutic aspect of the work that was supposed to be an important part of the office's task: The work with the individual client has largely been of a diagnostic and advisory nature, even in those cases where the professional staff knew that with longer efforts, they could really provide effective results.<sup>51</sup>

Several reports from Rakvåg stressed the urgent need to prioritize—that only children with pronounced and complex difficulties should be referred for individual assessment, and children from first to third grade were given precedence. An increase in nervous disturbances and adjustment difficulties was registered, and in the following years, the reports emphasized the significant number of adaptive difficulties [*tilpasningsvansker*]. In 1966, these made up 143 out of the 376 referrals.<sup>52</sup> Of these, 40 percent were considered very demanding cases that took up half of the work effort. The close relationship between adaptive difficulties and educational problems was pointed out several times. Reports also stressed that the shortage of suitable special educational and treatment measures made it difficult to follow up with complex individual cases:

---

50 Annual report 1962–63.

51 Annual report 1962–1963.

52 Annual report 1966–67.

'The work with these children is extremely demanding and constitutes a rather large additional burden for the professional staff, in that it leads to far too little of any positive results. The work seems pointless, partly because the possibilities for placement in observation school/class, special school, treatment home, foster home etc. are very small, partly because the office does not have the capacity to provide help and support over a longer period, in cases where it is believed that the difficulties could be solved within the child's usual home/school environment.'<sup>53</sup>

The reports from the 1960s are not very explicit on the use of specific assessment measures. They do, however, describe that students who were registered individually were tested by the school psychology office using personality and intelligence tests as one of several stages in the examination of the schoolchild's condition and needs. In 1968–69, regular visits were carried out to all schools, and 184 children were individually tested in 1968, compared to 162 children in 1969. Reports from the late 1960s still expressed concerns about the lack of staff, as well as a lack of measures aimed at children with mental and behavioural challenges.

Rakvåg stressed the need to limit the number of individual referrals through improved routines for school collaboration and teacher consultancy. A system with fixed attendance days at primary schools aimed at getting in early contact with teachers over challenging cases and avoiding too many formal requests for individual assessment was recommended. The reports from Rakvåg increasingly enhanced the need for new work methods, based on principles that resembled a "systemic" work model. One important argument was the resources demanded to carry out individual assessment:

'A school psychology examination is quite broad and time-consuming, with efforts from parents, children, teachers, doctors, special education teachers, social workers, psychologists and other staff. Such examinations should be reserved for relatively few children who have major difficulties. (...) By using the professional staff as consultants to the teachers and parents, you will reach far more children than with a traditional scheme, and this will perhaps be the most effective way of carrying out preventive work.'<sup>54</sup>

There was, however, another argument for keeping the waiting lists short. This was important not only to handle the workload, but to avoid

53 Annual report 1966–67.

54 Annual report 1969–70.

the harmful effects on children standing in line for help. Further, Rakvåg was explicit in praising his staff for not succumbing to a temptation to compromise with the requirement for professional and ethically sound work through accommodating the school's desire for more and faster investigations by 'more light-hearted and superficial work'.<sup>55</sup>

The office was responsible for statistical processing of pedagogical tests in mathematics and Norwegian. Such pedagogical tests were regarded as helpful when selecting pupils for further examination. Extensive reading and writing difficulties were described thoroughly in the reports, and linked to overcrowded schools, large classes, and time pressure.<sup>56</sup> However, the shortage of speech therapists was regarded as a barrier for the follow-up of such challenges. In 1966–67, a small municipal speech therapy office was incorporated and converted into a special education department that handled all cases of speech, reading, and writing difficulties. Facilitated by the employment of a social worker, the office also strengthened therapeutically-oriented work, focusing on follow-up of difficult cases via contact over time with children and their families. In 1969–70, the office reported complete staffing, with a team consisting of three psychologists, two social counsellors, two speech therapists, as well as a reading teacher, a primary school teacher, and a kindergarten teacher.

School preparedness tests were still carried out voluntarily during the 1960s, in collaboration with teachers and school doctors. Annual examinations of all school starters were requested by teachers as well as parents, but according to the annual reports, the EPS office lacked capacity for a comprehensive system of preparedness testing and had to prioritize accordingly.<sup>57</sup>

The late 1960s saw a change of prioritization and new ways of working in the Bergen EPS office, enhancing the collaborative and advisory work towards schoolteachers, and efforts to reduce the demand for individual assessment. The reports reflect changing moods towards the differentiated special needs educational system, and increased focus on integrating all children in local schools. The focus on individual assessment was reduced, and efforts were made to enhance preventive work and

---

55 Annual report 1968–69.

56 Annual report 1963–64.

57 Annual reports 1963–64/1964–65.

advising schoolteachers. Explicit critique of testing tools and practices is scarce in the material; in one of his last reports, however, Rakvåg stated that the EPS was too focused on repairment for children with severe problems, and expressed worries that differentiation based on criteria linked to deviation could have unfortunate consequences for the young children. He recommended enhancing what he labelled 'prophylactic work' through increased competence in normal pedagogy and normal psychology, and even stronger collaboration with the teachers.<sup>58</sup> The reports maintained that the office had increasingly prioritized individual and family treatment due to a lack of other local measures, but made possible through a high level of competence amongst staff.

## Integrating Social Engagement and Test Development

The national education law and the local framework for the EPS services changed during the late 1960s, together with the principle that most children should be educated in their local communities, and in 'ordinary' schools. While the Bergen office was not part of any official trials, the annual reports show that the local guidelines opened for new methods, with enhanced focus on more system-oriented work. According to the county development plan, the EPS should be a dynamic advisory body in a developing school system, and take part in pedagogical development efforts.<sup>59</sup> However, testing and advice regarding children who were considered to have severe learning difficulties or behavioural problems remained a crucial part of the work, according to the reports. From 1976, the principle of inclusive education became a national ambition through the new Education Act, which also expanded the user group.

On a local level, a municipal merger increased the geographical scope of the Bergen EPS services, and the annual reports changed form and were now written by the school psychologist from each district of the new and larger municipality. According to these reports, the collaborative relations between EPS and the local schools expanded, with more frequent school visits and systematic follow-up investigations

<sup>58</sup> Annual report 1971–72.

<sup>59</sup> Annual report 1967–68; Skoledirektøren i Bjørgvin, '*Utviklingsplan for spesialpedagogiske tiltak i Bergen og Hordaland*' (Bergen: J.W. Eides forlag, 1970).

where teachers reported on the effect of the measures.<sup>60</sup>

In 1971–72, 712 school children received service from the Bergen EPS. 398 of these children received assistance through collaboration between a teacher, psychologist, and social worker, without individual examination. For 314 children, it was regarded necessary to carry out individual examinations through psychological testing. One third of these children who were individually examined were referred through contact with the school psychology office, the rest directly from the schools. In 1976, the city centre district served 13,000 schoolchildren and 8,500 kindergarten children with the resource of 5¾ psychologist positions, 2 social workers, and ¼ speech therapists. Nonetheless, reports complained about understaffing. New positions had been approved, but hiring was delayed.

New priorities can be read from these annual reports. From 1976, the EPS was required to also serve pre-school children. Registration of individual pre-school children took place via contact with the kindergartens each spring. Earlier highly prioritized preparedness tests such as Sandven's test continued into the early 1970s, but were given less weight in the reports. Following the changes in the school act from 1976, attendance at the testing dropped as new models were tried out to prepare for the start of schooling.

Some of the early 1970s district reports detail their working methods. Regular school visits, conversations with class teachers, and classroom observation appear key. Individual assessment of abilities was mentioned as part of the work, but with less detail. More weight was given to the role of proficiency tests. One district report described the work routines thoroughly.<sup>61</sup> As an example, the EPS counselor Skiftun mentioned that he did not always meet the individual student, but met with parents or teachers. Only in more severe cases did individual student conversations and assessments take place, combined with parent interviews as well as individual therapy and teacher meetings.

Several of the psychologists connected the educational problems amongst schoolchildren to a growth in social problems.<sup>62</sup> One rich report

---

60 Annual report 1971–72.

61 Psychologist Skiftun in annual report 1974–75.

62 Annual report 1973–74.

from a suburban district office stressed the efforts of getting acquainted with the district schools and local living conditions, and described a working day based on collaboration, regular school visits, home visits, and individual assessment of children. The reports still mentioned limitations for carrying out individual examinations and treatment, and the pressure of high demand for service from the schools:

‘However, the danger of too much contact with the district and its institutions increases pressure for individual examinations and psychological treatment as staff discover that the psychologist is available (exists), and frustration and irritation when these wishes cannot be fulfilled’.<sup>63</sup>

This echoes earlier reports; the psychologist recommended the strengthening of district staff to include a full team, with a social worker and special education teacher in addition to the psychologist as the remedy to solve this dilemma.

Another report mentioned learning difficulties as the dominant issue—dyslexia in particular. Little assistance could, however, be offered by the EPS, because of vacant special education teacher positions. This psychologist framed the challenges thus: ‘What is most difficult about the job as a school psychologist is precisely having to function as a professional in a field one’s education does not cover.’<sup>64</sup>

In 1974–75, the annual report from an underprivileged district stressed that a surprising number of pupils had reading or writing difficulties as a primary or secondary problem. Challenging socio-economic conditions in parts of this district made the education problems a small part of the complex challenges faced by children that, according to the psychologist, could not be solved within the school system alone.<sup>65</sup> The Bergen school psychologists were concerned about the consequences of social problems and regarded social and political changes necessary to meet such challenges.

A new element in the annual reports from the 1970s is the engagement of EPS staff in research and development projects. These efforts to improve work methods embraced close collaboration with the emerging psychology

---

63 Psychologist Toril Havik, annual report 1973–74:15.

64 Skiftun in annual report 1974–75: 24.

65 Annual report 1974–75

research and clinics at the University of Bergen through competence building, student internships, and projects. The annual reports from 1973–74 and 1974–75 mention, for instance, collaboration with the neurologist Professor H. Kløve, and with Professor H. J. Gjessing, a central figure in shaping Norwegian special pedagogy and educational psychology.<sup>66</sup>

A few staff members were directly involved in developing test tools, for instance, through partaking in the Norwegian standardization of the Illinois Test of Psycholinguistic Abilities (the ITPA project), funded by the national research council NAVF. Others participated in projects involving collaboration with local schools. Psychologist Ellen Greve regularly monitored students and spent time on observation and individual investigations in so-called maturity classes,<sup>67</sup> as well as ordinary classrooms at Ny-Kronborg school, in an underprivileged area of the city. According to Greve, 'It quickly became clear that a certain mapping was necessary.'

This school subsequently initiated a new system for special needs teaching and involved the EPS as well as professor Gjessing. Based on the assessment of first graders with the Kuhlman-Andersen test, a new teaching plan was tried at the Ny-Kronborg project. The aim was to take care of the individual child, the class, and the teacher through focus on well-being, educational progress, and cooperation. Instead of sending individual students out of the classroom to be taught in special groups, the project tried out a two-teacher system. To meet the challenges of widespread social problems and lack of parental collaboration, the project involved newly engaged EPS social workers. The report claimed: 'At this school there is a crying need for curatorial help.' This local project thus made use of a model where testing was combined with trials to organize teaching and classrooms in new ways, with the aim of keeping all children in the same classroom. A social work approach supported these efforts, according to the project reports.

In the late 1970s, the EPS staff was also strongly involved in a large longitudinal study of reading and writing difficulties amongst schoolchildren, carried out in the period 1977–1987. The so-called

---

66 *Spesialpedagogikk - perspektiver. Festskrift til Professor Hans Jørgen Gjessings 70 års dag*, ed. by Terje Ogden and Ragnar Solheim (Oslo: Universitetsforlaget, 1990), p. 11. See also chapter 10 in this volume.

67 This was a municipal measure aimed at first graders who were considered too immature to start in ordinary classes.

Bergen Project was funded by the NAVF and the Nordic Council of Ministers [*Nordisk Ministerråd*]. It was led by professor Gjessing with the EPS office as partner, together with the Bergen school authorities.<sup>68</sup> In 1977, all Bergen municipality school starters were tested and described via social, pedagogical, and psychological criteria (3000 children). The study embraced classroom observations, teacher reports, assessment of student well-being, tests in language and math, as well as ability. The screening defined a narrow group of students with general and specific reading problems (approximately 273 students) whose cases were followed up by a range of experts. The individual screening through standardized tests and interviews with teachers and parents was carried out by EPS staff, together with other psychologists, pedagogues, and social workers. According to the annual report, one of the psychologists spent half of his working hours linked to the Bergen project in 1978–79, and the following year he became a member of the project's steering group. At that time, it was reported that the Bergen project had seized large parts of the office's capacity.

The work methods were based on county plans from 1970 and 1976 that strongly leaned towards the integration approach for special needs education, and continued to emphasize system-oriented work in schools in combination with various individual assessments. Throughout the 1970s, the role and methods of the EPS were debated. Critical voices, not least from the expanding social pedagogy profession, pointed at the inadequate frameworks that hindered new visions for educational integration.<sup>69</sup> The Bergen EPS reports in the 1970s often relate educational difficulties amongst children to socio- economic challenges. Through participation in several research projects, new methods for pedagogical development were tried out. These collaborations also contributed to the development of professor Gjessing's visions for the EPS and special pedagogy, combining broad individual assessment with educational and social measures.<sup>70</sup>

68 The Bergen project was prestigious and large, involving national and international partners. It was, however, not uncontroversial to EPS representatives. One of the few to raise critique towards the project approach and methods was Sigvald Asheim, who regarded Gjessing's work as a continuation of the diagnostic practices that had lost its legitimacy amongst many educators, see Dalen, *Integrering*, pp. 37–38.

69 One critical voice was special pedagogy teacher Eiliv Solum, who in 1981 declared that nothing had changed after the new act in 1976: Eiliv Solum, 'En kritisk analyse av den pedagogisk—psykologiske diagnostikk praksis i lys av skolereformene. 1980', *Skolepsykologi: Tidsskrift for pedagogisk- psykologisk tjeneste*, Vol. 16 no. 2 (1981), 3–12.

70 One of Gjessing's concerns was that many staff in different services, including the



## Discussion and Conclusion

Through a detailed scrutiny of the annual reports of one of the early EPS offices in Norway in the period 1953 to the late 1970s—with a particular focus on ambiguities and tensions regarding the role of psychometric testing—we have seen how the EPS, since its inception, has been regarded an important vehicle for the development of a comprehensive Norwegian education.<sup>71</sup> Some studies have regarded the shift from differentiated special needs education towards inclusive education a step towards new forms of social regulation.<sup>72</sup> Studies based on journals from the Norwegian EPS, however, underline the continuing, strong role played by individual IQ testing as a central feature of the services, despite education reforms, new approaches, and a redefining of the services.<sup>73</sup> This chapter has sought to explore how the testing practice was described and legitimized throughout a period with changing approaches towards the role of psychometric testing, and new frames for the services it included.

Acquirement and use of individual and group testing tools were a crucial part of how the work of the EPS office was defined throughout the period from 1954 until 1980. The diagnostic work embraced individual assessment of abilities, as well as personality and pedagogical testing. Assessments always combined various tests and involved a range of other methods, such as interviews and observations. IQ test results were never considered alone. The descriptions in the reports do, however, show shifting roles of different testing forms throughout the period.

The EPS offices were established to serve the differentiated special needs education system, and were built on models from the interwar period, with psychometric testing as a crucial element of the assessment, together with elements of observation and family work. The 1950s reports show that the heads of the new EPS office considered it important to underline how they put individual as well as group testing to the forefront. The reports stressed the efforts to equip the office with a range of available testing material of both international as well as Norwegian standardizations, and gave thorough

---

EPS, had received little formal training in using testing tools. His suggestions for improvement of testing tools aimed at solving diagnostic challenges in a situation with lack of educated staff and rising workload. (Gjessing 1974)

71 Froestad & Ravneberg, Education policy; Simonsen, Vitenskap.og profesjonskamp.

72 Hamre & Ydesen, *The Ascent*.

73 Tvedten, *Sakkyndighet*.

accounts of the usefulness of the different tools.

During the early 1960s, tasks were defined as mostly diagnostic, with reports underlining the limited resources to carry out therapeutic measures. Activities related to group school preparedness as well as pedagogical tests were increasingly significant. With growing workloads, methods were altered to prevent a rising number of referrals to individual assessment, giving priority to preventive measures and close collaboration with schools and teachers. During the late 1960s, certain critical views regarding testing can be traced. The descriptions also referred to new work models stressing the importance of more system-oriented work. Some reports from the late 1960s and early 1970s share concerns about possible negative effects of individual assessment, possibly leading to stigmatization of young people. Individual IQ and personality testing was, however, reported as important elements in handling difficult cases.

From the early 1970s, the services engaged in collaborative efforts with the expanding psychological research and clinics at the University of Bergen. EPS staff engaged in projects at local schools, together with professor Gjessing, who tried out his ideas of how to carry out counselling through a combination of standardized individual psychometric testing tools with observation, pedagogical testing, the addition of teacher resources, and advice and social work.

Throughout the whole period, the reports stress how the use of individual testing tools was restricted by access to competent staff, and time constraints. Tasks were solved in close collaboration with teachers, school authorities, and specialist health services. The reports highlight work pressure and demanding tasks, but also depict how staff actively tried to redefine and reorganize their activities in collaboration with municipal and county authorities.

These reports show how the roles of various assessment practices were given meaning by the leaders of a municipal service in a period where guidelines, tasks, as well as the values related to assessment changed. Rather than clear shifts regarding the use of individual testing, this analysis identifies a pragmatic and sometimes ambiguous approach. It is hard to trace clear disruptions in the approach to psychometric testing. The reports do not explicitly mention the national educational reforms, nor public debates on education and the role of assessment. Local guidelines are more present, but to a large degree seem to reflect the changes on a national level. The material, however, demonstrates

how new work principles and expertise add to existing practice. They give us a glimpse into how different tools for assessment and advice were used pragmatically to meet the changing and increasing demands from teachers and parents, and solve the tasks given by national and local regulations. According to Haug, the 1976 education reform implied a merging of two school laws.<sup>74</sup> Neither the principles of inclusive education nor the role of the EPS were clearly defined until the 1980s. Nevertheless, the focus of the 1970s reports on the social aspects of the learning and behavioural difficulties experienced by the school system might be interpreted to point to further development of services during the 1980s, when the EPS became more clearly defined as an expert organ of the education system.

## Bibliography

- Andreassen, Karen E. and Kristine Kousholt, 'Minorities and educational testing in schools in Arctic regions: An analysis and discussion focusing on normality, democracy, and inclusion for the cases of Greenland and the Swedish Sami schools', in *Testing and Inclusive Schooling: International Challenges and Opportunities*, ed. by Morin and Ydesen (London: Routledge, 2018).
- Årsmeldinger (annual reports), Skolepsykologisk kontor i Bergen, Bergen Byarkiv, Årsmeldinger, pedagogisk- psykologisk tjeneste, Sentrum. Arkivnr. A-3253, XA.
- Asdal, Kristin and Ingunn Moser, 'Experiments in Context and Contexting', *Science, Technology & Human Values*, 37.4 (2012), 291–306, <https://doi.org/10.1177/0162243912449749>
- Asdal, Kristin, 'What is the Issue? The Transformative Capacity of Documents', *Distinktion: Scandinavian Journal of Social Theory*, 16.1 (2015), 74–90, <https://doi.org/10.1080/1600910X.2015.1022194>
- Axelsson, Thom, 'Rätt elev i rätt klass. Skola, begåvning och styrning 1910–1950' (Doctoral dissertation, Linköpings universitet, 2007).
- Ball, Stephen J., *Foucault, Power and Education* (London: Routledge, 2013).
- Dalen, Monica, *Integrering av funksjonshemmede i grunnskolen- en prinsipiell drøfting* (Oslo: Universitetsforlaget, 1985).
- Evjen, Geir, *Pedagogisk- psykologisk rådgivning- et historisk perspektiv* (Oslo: Monography, Skolepsykologi, Materiellservice, 1984).

---

74 Haug, Ideal og realitet.

- Froestad, Jan and Bodil Ravneberg, 'Education Policy, the Norwegian Unitary School and the Social Construction of Disability', *Scandinavian Journal of History*, 31.2 (2006), 119–143, <https://doi.org/10.1080/03468750600604119>
- Gabrielsen, Egil and Kjersti Lundetræ, 'Indikerer de norske PIRLS-resultatene et behov for å justere retningslinjene for skolestartsalder?', in *Klar framgang. Leseferdighet på 4. og 5. trinn i et femtenårsperspektiv*, ed. by Egil Gabrielsen (Oslo: Universitetsforlaget 2017), pp. 204–221, <https://doi.org/10.18261/9788215030258-2017-12>
- Gjessing, Hans J., 'Norsk spesialundervisning gjennom tre decennier 1950–1980', *Spesialpedagogikk*, 3 (1972), 2–13.
- Gjessing, Hans J. (1974), 'Diagnostiske hjelpemidler i den pedagogisk-psykologiske rådgivningstjenesten, Innstilling fra utvalget for utredning av behovet for diagnostiske hjelpemidler (pedagogisk-psykologisk instrumenttjeneste), oppnevnt av Rådet for samfunnsvitenskap, NAVF), 8. Februar 1973, avlevert 13. Juni 1974.
- Hamre, Bjørn and Christian Ydesen, 'The Ascent of Educational Psychology in Denmark in the Interwar Years', *Nordic Journal of Educational History*, 1.2 (2014), 87–111, <https://doi.org/10.36368/NJEDH.V1I2.40>
- Hamre, Bjørn, Thom Axelsson and Kari Ludvigsen, 'Psychiatry in the Sorting of Schoolchildren in Scandinavia 1920–1950', *Paedagogica Historica*, 55.3 (2019), 391–415, <https://doi.org/10.1080/00309230.2019.1597137>
- Haug, Peder, 'Ideal og realitet i spesialundervisninga', in *Pedagogikkens idé og oppdrag*, ed. by Myklestad, Thuen and Vik (Bergen: Fagbokforlaget, 2022).
- Innst.O. XIV (1968–69). *Innstilling frå kirke- og undervisningskomiteen om lov om grunnskolen* (Ot.prp. nr. 59 for 1966–67).
- Kyrkje- og undervisningsdepartementet, St.meld. nr. 9 (1954) *Om tiltak til styrking av skoleverket*.
- Lindblad, Sverker, Daniel Pettersson and Thomas S. Popkewitz, eds., *Education by the Numbers and the Making of Society: The Expertise of International Assessments* (New York: Routledge, 2018), <https://doi.org/10.4324/9781315100432>
- Lov nr. 42 om brigde i lov av 13. juni 1969 nr. 24 om grunnskolen, datert 13. juni 1975.
- Ludvigsen, Kari and Åsmund Arup Seip, 'The Establishing of Norwegian Child Psychiatry: Ideas, Pioneers and Institutions', *History of Psychiatry*, 20.77 (2009), <https://doi.org/10.1177/0957154X08089847>
- NOU 1983:4, *Pedagogisk-psykologisk rådgivningstjeneste* (Flekkøyutvalget) (Oslo: Kirke- og undervisningsdepartementet).
- Ogden, Terje and Ragnar Solheim, eds., *Spesialpedagogikk- perspektiver. Festskrift til professor Hans Jørgen Gjessings 70-årsdag* (Oslo: Universitetsforlaget, 1990).

- Ot.prp. Nr. 4, 1955. Om brigde i lov om folkeskulen på landet, folkeskulen i kjøpstedene og lov om framhaldsskular (Oslo: Kyrkje- og undervisningsdepartementet).
- Reinertsen, Hilde and Kristin Asdal, 'Calculating the Blue Economy', *Journal of Cultural Economy*, 12.6 (2019), <https://doi.org/10.1080/17530350.2019.1639066>
- Rose, Nikolas, *Powers of Freedom: Reframing Political Thought* (Cambridge: Cambridge University Press, 1999), <https://doi.org/10.1017/CBO9780511488856>
- Samordningsnemnda for skoleverket, oppnevnt ved kongelig resolusjon 7. mars 1947, *Sammenfatning og utsyn*, (Mysen, Indre smaalenenes Trykkeri, 1952).
- Simonsen, Eva, *Vitenskap og profesjonskamp. Opplæring av døve og åndssvake i Norge 1881- 1963* (Oslo: Unipub forlag, 2000).
- Solum, Eiliv, 'En kritisk analyse av den pedagogisk–psykologiske diagnostikks praksis i lys av skolereformene', *Skolepsykologi*, 16.2 (1981), 3–12.
- Tveitnes, Marianne Sandvik, *Sakkyndighet med mål og mening: En analyse av sakkyndighetens institusjonaliserte kjennetegn; et grunnlag for refleksjon og endring* (Doctoral dissertation, University of Stavanger, Norway, 2018).
- Skoledirektøren i Bjørgvin, *Utviklingsplan for spesialpedagogiske tiltak i Bergen og Hordaland* (Bergen: J.W. Eides forlag, 1970).
- Vormeland, Oddvar, 'Pedagogisk-psykologisk rådgjeving i skolen' (Oslo: Fabritius, 1963).
- Ydesen, Christian and Karen E. Andreasen, 'Historical Roots of the Global Testing Culture in Education', *Nordic Studies in Education*, 40 (2020), 149–166, <https://doi.org/10.23865/nse.v40.2229>
- Ydesen, Christian, Kari Ludvigsen and Christian Lundahl, 'Creating an Educational Testing Profession in Norway, Sweden and Denmark, 1910–1960', *European Educational Research Journal*, 12.1 (2013), 120–138, <https://doi.org/10.2304/eerj.2013.12.1.120>

## 7. IQ Testing Today in Norway's Educational Psychological Services

*Håkon Aamot Caspersen*

---

From their early beginnings and rapid expansion at the start of the twentieth century, one of the main applications of IQ tests has been in different assessment procedures involving children. This is also the case in Norway, where educationalists and school psychiatrists, among others, spearheaded early adoptions and translations of the French Binet–Simon test, and its subsequent and more influential US revision, the Stanford–Binet.<sup>1</sup> The historical use of such tests laid the foundation for much of their continued institutionalization, such as within the Educational Psychological Services (EPS) [my translation of Norway's *Pedagogisk Psykologisk Tjeneste*], one of the key institutions where IQ tests are used today. This chapter provides some initial glimpses into their contemporary use to shed light on the continued influence of these technologies, their role and function within this disciplinary community of practice, and how IQ testing is understood, critiqued, and practiced. The chapter concludes with general reflections on these tests as professional markers and communicative devices.

This chapter is written by a trained anthropologist with one foot in science history and with one foot in the present. What follows is generally based on semi-structured interviews with budding, current, and former EPS counselors, most of whom belong to the central east of Norway. It also builds on more informal engagements with different

---

1 See Lefkaditou, chapter 1, and Ydesen, Hovland and Vikström, chapter 2, this volume. For a concise historical overview of the development of intelligence testing internationally, see Annette Mülberger Rogele, 'Biographies of a scientific subject: The intelligence test' (2020); see also Porter, chapter 13, this volume.

counselors on the subject of testing, in forums such as furthering education courses on the Wechsler Intelligence Scale for Children—version five (WISC-V) and an EPS network meeting on the assessment of children with minority backgrounds where testing was discussed, to mention a few concrete examples. This is part of ongoing research that includes ethnographically informed classroom observation of taught components across different university-based test-certification modules, in combination with interviews with some of the students. While I am broadly familiar with the work of the EPS and tests such as WISC—their administration, scoring, analysis, and how they are taught—the focus remains squarely on those conducting the tests, not on those being tested. Due to practical and ethical reasons involving restricted access and confidentiality, I do not have direct experiences of testing in a real-life test situation, nor with the subsequent interpretation of the results in an expert assessment procedure. As a recent review of state-of-the-art research on EPS in Norway makes clear, there is an urgent need for more research into this lacuna, both ethnographic and otherwise, towards which this chapter is a first step.<sup>2</sup>

A further qualification on whose “voice” is found below is in order. The emphasis is on critical perspectives to testing. The reason is twofold. Firstly, it has to do with the group of informants who have agreed to share their reflections on testing, the majority of whom were of a critical inclination—or, at least, what shone through their interviews was that tests, as such, were understood as imperfect, but nevertheless served important functions that could help when in the hands of the right experts. When it comes to the “representativeness” of different opinions on testing, one must keep in mind the observation of two psychologists working at the EPS in Norway’s largest city: there will be people that are very critical of testing, and a few that are very positive about testing, but most will be somewhere in the middle, critically considering it a practical tool to gather information as part of a larger assessment procedure. Secondly, with the above in mind, criticism and opposition have the potential to make visible internal disciplinary debates, critical reflections, practices, and assumptions that normally might be hidden from view, which, incidentally, is a heuristic fit with our purpose

---

2 Toril Moen et al, ‘The Norwegian Educational Psychological Service’, *Nordic Studies in Education*, 38.2 (2018), <https://doi.org/10.18261/issn.1891-5949-2018-02-02>

of providing a glimpse into how IQ testing is viewed by some of its practitioners today.

IQ tests and IQ testing here are mainly represented by use of the Wechsler Intelligence Scale for Children, version V, colloquially known by the abbreviation WISC-V. Nationwide, 8 percent of children in primary and middle education received special education in the school year 2021–2022, equalling 49,000 pupils, with a gradual growth in cases until middle school. Conservatively, we can assume that more than half of these children have been subjected to the WISC, while a further one-fourth is very likely subject to other tests of general ability, such as Raven, Leiter, or CAS.<sup>3</sup> Information from informal interviews gives estimates in the range from one-third to two-thirds of cases when it comes to the use of a WISC. IQ testing, in short, is an everyday occurrence, institutionally integrated in the educational support system today, albeit with regional differences.

While I have often heard of the existence of “die-hard” supporters of tests, “essentialists”, or of people that treat IQ tests and their results as valid unchanging measures of biologically defined cognitive resources that one has more or less of, I have so far never met them among those who actually use these tests. There is a tension and a double nature to how these tests are employed and perceived in the EPS, where they serve both as dynamic assessment tools and as a baseline for further assessment. Testing is today perceived in explicit opposition to stereotypical notion of testing as an exercise in providing oppressive stamps of categorization, something that internally is talked about (both today, and historically) as malpractice resulting from an over-reliance on test results as a simple sorting mechanism.<sup>4</sup> Nevertheless, the tests continue to exert legitimacy on expert assessment procedures, imbuing the process with a type of neutrality associated with such tests and other scientific measuring instruments.<sup>5</sup> IQ testing here becomes a kind

3 Marianne Sandvik Tveitnes, *Sakkyndighet med mål og mening: En analyse av sakkyndighetens institusjonaliserte kjennetegn; et grunnlag for refleksjon og endring* (PhD thesis, University of Stavanger, 2018).

4 For a historic example of how such a critique has always been “historicized”, see for instance Hans Jørgen Gjessing et al, *Diagnostiske hjelpemidler i den pedagogisk-psykologiske rådgivningstjenesten: instilling*—a report from a committee formed to assess the need for diagnostic tools in the EPS. Published by Rådet for Samfunnsvitenskapelig Forskning, NAVF (1974).

5 Marianne Tveitnes (2018); Joron Phil *Etnisk mangfold i skolen: Det sakkyndige blikk*



of ritual—as a technology, it needs to be performed in a standardized way, depending on protected expert knowledge to produce the right effects (a valid score). This striving to achieve the standards of the tests, and the accompanying ideals of objectivity and validity, are important; but in an assessment process, the complexity of cases, institutional positions, personal preference, and the background of expert advisors shape and influence what they see and how they use and conceptualize tests, despite best efforts to exclude subjectivity from the equation.<sup>6</sup>

### WISC, an IQ Test Prototype

WISC first saw the light in 1949. As part of the larger family of Wechsler tests, it grew out of the original Wechsler-Bellevue developed in the late 1930s, with the manual *The Measurement of Adult Intelligence* first published in 1939.<sup>7</sup> It is now in its fifth revision, with the last translated and standardized Norwegian version appearing in 2017. Internationally, as well as in Norway, it is the most frequently used test of general ability, together with the Wechsler Adult Intelligence Scale. The test provides a full-scale IQ (Intelligence Quotient) score and a breakdown of performance according to five different factors or indexes, ‘areas of abilities’, that can be topographically visualized as points in a profile graph for analysis.

The test battery acts as a prototype, a “best example” or a gold standard of the state-of-the-art of psychometric tests of general intelligence today. A comparison of educational programmes in psychology, school psychology, and special needs pedagogy—which include a general certification of test competence in Norway—show that when students are taught psychometric testing, they learn in large part through a hands-on practical introduction to WISC with the aim of transferring the skills learned to other test batteries. WISC, it turns out, is very much thought of as a key tool of the trade.<sup>8</sup>

---

(Oslo: Universitetsforlaget, 2010).

6 Maarten Derksen, ‘Discipline, subjectivity and personality: an analysis of the manuals of four psychological tests’, *History of the Human Sciences*, 14 (1) (2001), 25–47.

7 David Wechsler, *The Measurement of Adult Intelligence* (Williams & Wilkins Co., 1939).

8 This does not mean that there are no existing alternatives in Norway. Although with a much smaller market share, companies like Hogrefe and Pedverket provide certification in tests that serve similar purposes to the WISC, often aimed at existing

Not everyone can use an IQ test as indicated. Both the testing itself, and who can test IQs, is a regulated activity that rests on access to protected discipline-specific expertise and to the test kits themselves, just as they have historically.<sup>9</sup> The rationale, as seen from the test field, continues to be connected to: 1) the powerful potential social and institutional effects of a score, and hence the importance of correct administration, analysis, and translation of the results to assure their validity; and 2) the importance of protecting the content of the test from the public, to ensure their integrity as valid measures of intelligence (to avoid “teaching to the test”).

In Norway, Pearson Clinical (part of the international media and publishing company Pearson Inc.), who are the owners of tests such as WISC-V, restrict its purchase to authorized psychologists, graduates of one MA programme in special needs education, and one MA programme in pedagogical-psychological advising. In addition, it is open to graduates of other MA programmes of pedagogy and special needs education who can document two years of practical experience from the pedagogical-psychological services, together with the completion of a named fifteen-credit university course in ‘Test-theoretical competence’ offered at a specific Norwegian university.

In practice, if we continue with WISC-V as an example, there are more informal routes to test certification, or access to the practical use of tests, that are more or less condoned, at least when it comes to testing within the EPS. In my engagement with EPS advisors, for instance, in the context of popular, independent, shorter WISC-V training seminars (the largest I attended had up to eighty participants), I met several newly employed EPS advisors on alternative paths to WISC certification.<sup>10</sup> Certification, in some cases, can take the form of supervised practice-based training internally, supplemented by or in combination with more formalized short-term, seminar-based

---

or new employees in, for instance, the educational psychological services. This is not to mention the wealth of other tests and recording forms available in the arsenal of the EPS; the availability and employment of such tests has been important from their early beginning, as amongst others Ludvigsen provide evidence for in chapter 6, this volume.

9 As was the case, for instance, with the Norwegian version of the Terman–Merill from 1954; see Paulsen, chapter 9, this volume.

10 See also Fay Erle Fjeld, *Bruk av WISC-IV i pedagogisk psykologisk tjeneste: Hvordan vurderer ansatte i PPT WISC-IV som et verktøy i utredning av barn?* (Master’s thesis, Universitetet i Oslo, 2015), pp. 64–65.

courses on WISC-V administration and interpretation. These more informal practices are perhaps propagated because it is usually institutions or their representatives that will pose as buyers of such test kits. What happens once they are applied in-house is more difficult to control.<sup>11</sup>

## IQ in a Qualitative Cloak

The EPS counselors I have so far spoken with do not talk that much about intelligence or IQ, but they do talk a lot about ‘*forutsetninger for læring*’ [capacities for learning]—phrasing based on a consciousness of how an IQ measure offers unfortunate connotations of finality and concreteness. This is also reflected in test manuals that advise using a set of standard qualitative descriptors when communicating the results to parents and clients, such as ‘clearly above average’, ‘lower part of average’, or ‘significantly below average’ [my translation of the Norwegian qualitative descriptions found in the most recent version of WISC]. Students of psychology and special needs pedagogy learn to be careful of mentioning an IQ score when they are trained in testing, as “people easily misunderstand” scores.

This also means that one might not generally hear the terms ‘IQ score’ or ‘IQ test’ mentioned when IQ tests are used today. The children tested are, of course, aware that they have been tested, but the test might be presented as one of general abilities, or as a set of tasks that assess how they learn, or even one that maps a general cognitive profile of learning capacities. But WISC remains a test of general intelligence, and results in IQ scores. Importantly, it also results in other information that is deemed of equal importance. As these technologies travel across time, place, and contexts, there might be changes in their names, people’s understanding, and even the theoretical foundations of tests, but other aspects remain the same.

---

11 A recent diagnosis of the current state of the field of the use of psychological tests in Norway, conducted as a collaboration between the Norwegian Association of Psychology [*Psykologiforening*], the health directorate [*helsedirektoratet*] and the Norwegian Institute for Public Health [*Folkehelseinstituttet*], supports such an assumption, describing a field characterized by a lack of centralized guidelines, and the fact that the number of the tests used by Norwegian psychologists are not necessarily controlled in terms of official translations and its psychometric properties. See for instance T. M. Ryder, ‘Testkvalitetsprosjektet—del 1: Norske psykologers testholdninger og testbruk’, *Tidsskrift for Norsk Psykologiforening*, 58 (1) (2021), and T. M. Ryder, ‘Testkvalitetsprosjektet—del 2: Tester med behov for kvalitetstiltak’, *Tidsskrift for Norsk Psykologiforening*, 58 (2), (2021).

## How Is a WISC Understood and Talked About by Those Who Use It?

To aid us in answering this question, we will hear from Anna and Petra—two MA students in special pedagogy who demonstrate how they, as future test practitioners who have recently completed an obligatory test certification module, reflect on these tests and their role and function in an assessment procedure. While far from all graduates are interested in a career in the EPS, their successful completion of the MA programme and the WISC-V component of the course contributes to a general test competence that will make them very attractive candidates for employment. While psychologists still very much “own” these tests, there is a disciplinary hierarchy—it is usually psychologists, for instance, that provide instruction and teach testing. But in Norway, in comparison to its Danish neighbour, for instance, they are not the only ones who conduct the tests. Despite the name ‘pedagogical-psychological services’, in direct translation, the composition of expertise today is perhaps better characterized, in the words of one EPS advisor, as a ‘pedagogical-special pedagogical service’. The number of employees with an education in psychology amounts to roughly ten percent when including all levels of education in psychology. If one introduces the criteria of qualification to be that of an authorized psychologist with further specialization that requires particular educational requirements and practical experience, estimates leave only one percent of employees.<sup>12</sup>

Anna and Petra’s reflections are drawn upon here because they explicitly see the MA as an entry point into a future career in the EPS. They are both in their late twenties, with backgrounds as a primary teacher and kindergarten teacher respectively. They have both volunteered as research assistants, through which they have gained further experience using and administering different tests such as WPPSI (another Wechsler test aimed at preschool children), either as part of the student clinic at the local test lab, or as part of a research project. While Anna specializes in speech therapy—her secondary choice, as she did not receive a place in the psychosocial difficulties specialization—Petra specializes in *utviklingshemming* [developmental disability].

12 T. M. Andrews and B. Hustad, *Bemanning I PP-Tjenesten: Tilstrekkelig for å oppfylle tjenestens mandat?* Nordlandsforskning (NF rapport nr.: 12/2022), pp. 51.

For them, WISC-V is a test that:

Measures a type of intelligence [...] But I do not think that high intelligence is the same as being smart, necessarily. A human being with a score of a hundred could very well be perceived as smart or intelligent because one possesses competences that the society values, but that the test does not capture. [...] If someone asks, I say that it is a test where one does different tasks; that it consists of fifteen subtests that connect to different indexes of abilities such as seeing connections, verbal explanations, visuospatial aspects [where everybody goes: what?...] [...] I usually explain this with how the brain receives information in different pieces, and that it is about how one puts all those sensory pieces together again [...] then it is about logical thinking, how well you tackle a situation based on what you know from before, and working memory and processing speed—how much you manage to remember at the same time and how quick things go. —Petra

In the above, we can also sense how the students, through their introduction and training, have rehearsed their presentation of the test, which they will bring with them into future professional settings. We also see how such a rehearsal is necessary, as the specialized structure and language of such tests are full of disciplinary jargon that needs translation, including for themselves and for this anthropologist.

Anna defined WISC as '*et måleinstrument for evner* [a measuring instrument of abilities] that is there to give children who need it support in normal education'. She went on to highlight her previous conceptions of intelligence as something 'one can train and change'. Now she sees 'FSIQ (Full Scale Intelligence Quotient) as giving a good picture of a child's general level of abilities' and as a more stable measure. She makes reference to research showing the stability of the test across time, and clarifies what she means through differentiation in what the test measures: 'Training and learning is indicated on tests of crystalized intelligence [...] but this is not the same as fluid intelligence, that is also measured in these tests'. According to Anna, to speak of these tests as measuring intelligence, but without understanding it as a concept that encompasses different aspects, is a central problem in relation to understanding intelligence as a phenomenon, and in explaining it to others. But in her opinion, an IQ score in and of itself is not the most relevant aspect of these tests:

One would need to interpret the result together with the profile, and primary and secondary analysis [...] it is all the parts together that are useful. While FSIQ gives you a comparison with other children, primary analysis gives you a comparison with yourself. This latter is very important in relation to when one is to think of special pedagogical measures, I think. —Anna

She added that ‘intelligence in Europe does not need to be the same as in Africa, because of differences in cultural adaptation’, and that the employment of WISC to map the abilities of minority children will give you a ‘wrong picture of that child’s abilities’.

Petra, alongside the other students, is careful to show a critical awareness, for example in relation to how these tests are normed for specific groups of people, and the problems involved in testing children with a minority background. However, Petra would not necessarily subscribe to a ‘snapshot perspective’ [sic] of these tests (as some of the EPS advisors I have met emphasized, where results are understood to primarily indicate performance at that moment in time):

When children are getting older, let us say the ages of nine and twelve, research shows correlations of, I think it is, 0.8 between results [...] If one is not onto something with WISC, there is not much point to it [...] I saw that working at the [student] clinic [...] Parents were worried about skills in mathematics and Norwegian, and I did like ten different tests connected to learning goals in the school—but that tells me nothing other than the learning goals of a Norwegian school. But WISC is not only about skills in books [...] it is about the longer lines and skills that one is using in life that always will be part of affecting how well you do in life. What tools you can pick out of your toolbox, sort of.

Many MA students might go into working life on the receiving end of test results, working on translating assessments into concrete pedagogical measures in schools. Petra is explicit about what she wants to test:

I want to test the children that need it. And I want to do it correctly. It is a working tool that we should respect, and not take too lightly [...] I am... not a statistician or a neuropsychologist, I don’t need to understand everything, but I need to know how to use the tool... follow the manual... know that it has a scientific base, and that A leads to B.

We have seen the general agreement that WISC should be used only in cases where there is a strong suspicion of a problem. But when Petra

thinks back on her training—despite appreciating that testing can only be one piece of the puzzle, that it should never be used alone, and that one needs to adjust the assessment procedures to the case at hand—she cannot help but also point out that ‘there is a sense that we here are raised into a way of thinking of WISC as the solution, of WISC as a measurement of truth’. She would envisage that ‘the use of WISC, in relation to the teaching received, is that one almost would use it every time. Like a start.’ She also nods to her experiences as a research assistant, and how she has seen testing practiced more flexibly than what is ‘preached in class’, based on other criteria based on the professional judgments of the psychologist.

That testing in practice might harbour differences compared to the approach taken in training is perhaps not surprising. One of Petra’s fellow students who had been on a training placement in the EPS, for instance, pointed out that one of the first things they learned was to ‘forget everything you have learnt about WISC’, with a critical skepticism that she found surprising.

### Interpretation: the Challenge of Testing ‘For Real’ and Learning on the Job

The biggest challenge of learning to use the WISC, the students agree, comes from learning how to correctly interpret and analyze results. This is experienced as a different type of knowledge than that of needing to develop fluency in practical administration.

Testing in theory and testing “for real” are different, as articulated by Helga, an EPS counselor with several years’ experience who graduated with an MA in special needs pedagogy ten years ago:

While we got a lot of depth connected to general principles and themes, when I started in the educational psychological services, it was only WISC I knew... other types of assessment tools I needed to figure out myself. You work with complex cases, where it often is very difficult to actually administer a WISC test. Because, of course, it should be administered in a very specific way. In some cases, this would translate to a more dynamic testing and meta reflections, and I felt I learnt just as much from this as the test result. Imagine if we would have learnt more about this? It is as if the test here becomes a method, sort of, as opposed to an instrument.  
—Helga

Helga's reflections point to the use of WISC as part of a larger whole, in the sense of being one element in an overarching assessment process. That said, most offices will have in place some kind of mentoring for new staff, but what form this might take has large regional variations.

When it comes to the expectations of how testing would play out in the EPS, the above-mentioned students cited previous individual experiences of the EPS as teachers and kindergarten teachers. The service would often not accept a referral without the school having already done initial assessments or tests with mapping tools that the EPS recommended as a first step, sometimes independent of the reason for referral. This led to an attitude of doing tests on a routine basis just to 'have them done', an impression partly confirmed in informal interviews with current and former EPS advisors as being common practice in certain regional offices. Such initial test results become an important background from which to start an assessment process, which is also the reasoning of the youth and child psychiatric services for wanting the EPS to conduct the WISC, so that they already have the 'blood work' [sic] done, as a baseline for when, or if, cases enter the psychiatric services for further assessment.

## Testing in the EPS

The EPS is a nationwide service organized by county and municipality, with corresponding regional differences. In its contemporary institutional guise, it can be said to have existed since the first offices for school psychological counseling in Aker and Oslo were created in the 1940s.<sup>13</sup> Today, it is a large, central, and integrated institutional service in Norway's educational system, with around roughly 2,100 employees distributed across just under 300 regional offices.<sup>14</sup>

The mandate of the service is rooted in education law.<sup>15</sup> Here, its

13 For a historical overview of the service's development, see Asbjørn Birkemo, *Pedagogisk-psykologisk rådgivning* (Unipub: Universitetet i Oslo, 2009). See also Ludvigsen, chapter 6, this volume for a more detailed examination of test practice in the EPS from the 1960s up until the 1980s).

14 Andrews and Hustad 2022, pp. 29.

15 The following was written prior to the implementation of the new law of education in 2024. In general it outlines overarching continuities in the work of the EPS, both in its structural advisory role, and as expert assessors—although their expert assessments are no longer directed to assess the need for special needs education, wording that



main tasks are defined as, on the one hand, furthering competence and organizational development in schools when it comes to the education of pupils with special needs, and the individual assessment of the need for special education on the other.

The general justification for the work of the EPS has strong historical continuities. One such continuity is the employment of tests of general abilities as central tools in assessment procedures. This does not mean that the practice has not changed over time in line with changes in the educational system. One oft-mentioned example is a shift from, for instance, the “main” function of the service as “separating out” pupils for special education classes, so that their assessment work now leads to pedagogical differentiation and special educational efforts in normal education. This is a shift described as having occurred in the 1970s, in line with the political and institutional implementation of integrative education.<sup>16</sup> Although, as Ludvigsen (chapter 6, this volume) shows, this change might have a more complicated trajectory.

It has often been argued that such developments saw corresponding shifts in the work of the EPS, from more individualized to systemic advisory work—although the conservative effects of practice-as-usual and expectations from schools made it difficult to assess how this affected work across the services.<sup>17</sup> Similar tensions are repeatedly reported as an issue today. In textbooks and in the lecture rooms, there are continued discussions about the importance of moving away from individual to more systemic approaches, despite the social-pedagogical models that came to the fore in the 1970s now dominating the services.<sup>18</sup> It is the pupils who, in their specific context, should be the object of diagnostic work and not the individual in relation to a norm. In practice, testing remains an important and institutionalized part of the work of the EPS, and it is assessment directed at individuals, not systemic work directed at advising on and implementing changes on the institutional level in schools, that takes up most of the time according to the advisors

---

has been removed, but rather to assess the rights to and give advice on the need of ‘individually adapted learning measures’ [*individuell tilrettelagt opplæring*], the need for personal assistance, and physical adaptations of the learning environment.

16 Birkemo 2009.

17 Ivar Bråten and Arne Kristian Henriksen, *Kvalifikasjonsproblem i tilknytning til den pedagogisk-psykologiske rådgivningstjenesten: Om utdanning av pedagogisk-psykologiske rådgivere* (Oslo: Seksjon for skolepsykologi, 1983).

18 As discussed by Bråten and Henriksen.

I have spoken to. Others have estimated the time spent on individual assessment and the writing of reports to account for up to eighty percent of total work time.<sup>19</sup>

This work includes a range of tasks and activities, from participating in school meetings, individual conversations and class observations, initial assessment practices using questionnaires, testing and interpretation, the writing of individual *sakkyndig vurdering* [expert assessments], and corresponding feedback to schools, pupils, and guardians.

Testing, in this regard, includes not only tests of general ability like the WISC, Leiter, or Raven, but also more specific tests and mapping tools that deal with language and reading sufficiency, attention, and mathematical understanding, depending on the case at hand. EPS advisors report access to a large array of tests, like Vineland-3, Bayley-III, Leiter-R and Leiter-3, Raven's 2, Vålertesten, LOGOS, CMAT, CELF-4, and Brief-2, to mention a few. These range from tests of adaptive skills and tests of language to mathematical ability, communication abilities, specific learning difficulties, and tests of general intelligence.

While there are different practices depending on the offices involved, assessment procedures are largely an individual task. However, many report group meetings where, among other issues, the interpretation of test results is regularly discussed. While testing is done on a regular basis, several of the EPS councilors I have met have nevertheless often pointed to a knowledge gap and lack of familiarity in relation to what these tests are and their theoretical foundation. While the part of the manual describing directions for administration is in frequent use, the parts on test norms and theoretical foundations are left at the office and tend to collect dust, and, in practice, scoring and analysis might be aided by computerized software programs. This is not a trivial issue, and previous studies have pointed to recent historic discrepancies between reported and actual practice in the EPS when it comes to the correct assessment of minority students, for instance. In not sufficiently taking into account the language factor when using these tests, minority students are assessed using tests developed for the majority, as if they were majority students.<sup>20</sup>

19 Thomas Nordahl et al, *Inkluderende Fellesskap for barn og unge: Ekspertgruppen for barn og unge med behov for særskilt tilrettelegging* (Fagbokforlaget, 2018), pp. 170.

20 Knut Erik Aagaard. *Den språklige faktor: Pedagogisk-psykologisk utredning av barn med*

One of the prevalent reasons for referral to the EPS today continues to be a concern, often stemming from schools, that the child does not reach developmental milestones, broadly speaking, at the expected time or age as their peers.<sup>21</sup> This is similar to the reasoning behind the first standardized editions of the Binet–Simon tests introduced more than a hundred years ago, employed as indicators of the likelihood that the child will be able to follow normal pedagogical instruction in their age group based on their performance on the test, in relation to the expected normal results amongst peers. This was a test originally scored in the form of a child’s mental age—the predecessor of IQ scores today, explored in other contributions in this volume in more detail.

Even today, although in rare cases, the resulting IQ score might be converted and translated for purposes of special pedagogical intervention as a type of mental age, where in relation to finding the level of concrete tasks and appropriate pedagogical measures, it might be useful to “think with” the information that a ten-year-old performs on a level normal for the age group of eight years, for the sake of analysis. That said, in most cases, one might not directly report a numerical IQ score or mental age when dealing with schools and parents (although some do). A few of the EPS advisors I spoke with had experiences with schools that tend to ‘hang themselves up, in and in a way become very focused on it’, to quote Augusta, an EPS advisor based in a suburban area of Norway’s largest city. In cases where concrete IQ scores have become known to schools, another EPS advisor reported that there could be a certain resigned attitude among teachers, for instance.

The position implicitly critiqued above is that of an essentialist view of intelligence, which treats it as an interiorized, stable, natural capacity that one is born with, and the idea that this can be correctly represented and quantified through testing in the form of an IQ score. In the EPS advisors’ perspectives, this ignores a more proper and dynamic interpretation of the result, while recognizing that ‘there is something to the score’, and that, for most, a score seen as an indication of current performance under certain specific conditions. Expert evaluations might be revised as years pass; for

---

*minoritetsspråklig bakgrunn* (PhD thesis, University of Oslo, 2011); Jorun Phil, *Etnisk mangfold i skolen: Det sakkyndige blikket* 2<sup>nd</sup> ed. (Universitetsforlaget, 2010). See also Kyllingstad, chapter 10, this volume.

21 Finn Hesselberg and Stephen von Tetzchner, *Pedagogisk-psykologisk arbeid* (Gyldendal, 2016).

instance, I have seen an example where the positive discrepancy between the first WISC score and the second WISC tests a couple of years later was used to argue for the effect of the special needs educational measures received. So, an IQ score is understood by many as open to change and development, and not as a total defining measure. As a statistical measure, the score is never “the score”, only a likely estimate, based on a confidence interval usually set at ninety-five percent, translating to an estimate within five points in either direction on the scale.

While the total score is always reported (often in qualitative terms), what is described in more detail in reports are the different indexes, or cognitive profiles.<sup>22</sup> Expert assessments generally discuss the results of WISC at some length, describing as the reason a need to ‘translate it’ and make it useful for the schools that might not have the necessary background to interpret the findings without this type of contextualization.

## Tests Do Not Think for Themselves

In a lecture on the theory of intelligence, a quote from a review article on psychological testing and assessment was employed to emphasize this message:

Tests do not think for themselves, nor do they directly communicate with patients. Like a stethoscope, a blood pressure gauge, or an MRI scan, a psychological test is a dumb tool, and the worth of the tool cannot be separated from the sophistication of the clinician who draws inferences from it and then communicates with patients and other professionals.<sup>23</sup>

The ability of professionals in the EPS to draw inferences and make good use of a testing situation seems to not be limited to the ability to interpret and analyze the result in relation to performance within a normed group as found in the manuals. Of equal, and, in certain cases, heightened value is how testing with a standardized test of general intelligence allows them to draw a baseline to compare the performance to other cases that they have personally observed and assessed. My impression is that personal

---

22 Exemplified in some detail in Tveitnes 2018.

23 Gregory J. Meyer et al, ‘Psychological Testing and Psychological Assessment: A Review of Evidence and Issues.’, *American Psychologist*, 56.2 (2001), 128, <https://doi.org/10.1037/0003-066X.56.2.128>

experience with testing is seen as very important in this regard, and it is on this basis that EPS advisors sometimes will argue for the validity of test results and the use of tests that might not be normed for that group. A case in point is the frequent referrals to the EPS of children from minority language backgrounds with little or basic competence in Norwegian. In an informal discussion with several EPS advisors, Katja, an EPS counselor in a larger town in the middle of Norway, pointed out that:

The largest and most frequent grounds for referrals when it comes to pupils with a multi-lingual background from the side of the schools I work with, is suspicion of faults in cognitive functioning. Often together with psychosocial difficulties... here one of the main tasks with, for example using Leiter [a non-verbal test of general intelligence; she does not use WISC as in most cases the pupils' language skills are not developed enough to make testing feasible] and dynamic testing, is to disprove this—to show that this is about language, not cognitive development.

Katja, during our conversation, explicitly pointed to a dynamic approach to testing and mentioned how she finds Leiter to be a test that lends itself well to creative use. Based on her own extensive experience with using this test, she finds that a pupil's performance on the different tasks allows her to quickly form a good foundational understanding of the underlying issues at hand, in an explicit and implicit relation to her own personal experience. The mention of Leiter-3 in this instance is also interesting in that this is a test that, in opposition to WISC, is not normed for the Norwegian population, although it is translated into Norwegian. The prevalence of such tests in situations to which the test is not adapted has a long historical precedent in Norway. In a recent assessment of the situation, psychologists described this 'muddy field' under the heading 'a test field out of control'.<sup>24</sup> While being aware of this issue, and of problems connected to the cultural bias of tests, Katja circumvents this 'inherent weakness', in her opinion, through reliance on her own professional judgment, based on personal comparative experience of using this test on many different occasions on

---

24 Øystein Helmikstøl and Nora Skjerdingsstad, 'Testfelt uten styring', *Tidsskrift for Norsk psykologforening*, 58.1 (2021), <https://psykologtidsskriftet.no/nyheter/2021/01/testfelt-uten-styring>; Tanya Marie Ryder, 'Testkvalitetsprosjektet - del 1: Norske psykologers testholdninger og testbruk', *Tidsskrift for Norsk psykologforening*, 58.1 (2021), <https://psykologtidsskriftet.no/rapport/2021/01/testkvalitetsprosjektet-del-1-norske-psykologers-testholdninger-og-testbruk>

children with similar or different problems (in combination with other tests and information in expert assessment procedures). She also mentions that the pupils with a multilingual background referred to the EPS would not usually fit into a Norwegian normed group in any case.

### Testing in the EPS Today: Critical Distance and Different Practices

Why would you use an IQ test or test of general ability in the PP-services today? Some would put it this way:

“You test because this is what you should do. You test because one shall get a better impression of a child’s *læreforutsetninger* [capacities for learning]. Or, to see whether there is reason for a further referral to the child and youth psychiatry, or wherever...”

‘You get a pupil that is very anxious, the response from one’s surroundings is: “Have you done a WISC?” You get a pupil with a very specific and clear difficulty in reading and writing: “Have you taken a WISC?” If we got a pupil who had been in the system before, but who had not been tested, then “he had to be tested” ...’

““Let us take a test” in the sense of doing something with or dealing with the issue properly... [The test] is something important that will solve something... a key, sort of.’

‘[Testing] is, in a way, practice as usual. But there are no forms or templates that tell us this. There is no handbook for EPS advisors, where it says “remember to take a WISC”, at least I have not gotten any.’

—Quotes compiled from conversations with three advisors in the EPS that were working or had worked in an urban area in Norway’s central eastern region.

The above exchanges are clearly critical, all from those educated as pedagogues, recruited to undertake back-to-back interviews, recommended based on having “interesting views” that often translated as a critical perspective on testing. The main message is that there is too much testing, or not enough correct testing; too little critical reflection on testing as it is usually practiced; and that there is a test optimism that stretches beyond the EPS, where taking a test becomes synonymous with getting something done properly.

Testing, in general, seems to be on the agenda and is something that is critically and frequently discussed in the EPS. This includes a plethora of general and specialized tests, where intelligence tests such as WISC only become part of a wider picture—if not of testing itself, then at least of how it is practiced and discussed. Here, for example, suggestions of initiatives to increase advisors' skill and knowledge of testing are often made, as are efforts to make test praxis more standardized. When pushed to be specific about their critiques, the interviewees tend to mention examples on a more generalized level, perhaps referencing a colleague that, in their opinion, 'tests too much' or 'she really likes testing and always brings a WISC'. In the instance quoted here, Moira, an EPS advisor from a smaller rural counseling office, quickly added, 'But I must say, she is very effective and gets through a lot of cases, and it is not like I think that she is not doing a good job'.

Even such internal critique recognizes testing as an important, albeit far from perfect, tool of the trade. It is the best tool they have to do the job needed, subject to professional expertise, judgment, and intuition, to ensure best practice and correct use based on their knowledge of what these tests are, what they measure, and when they should be used.

The reason for the continued priority of testing that the interviewees described in their own work, despite this critical attitude, was the high level of pressure at work and limited time to complete evaluative assessment procedures. Testing, not least the meeting and the information to be gained through the observation of the child in a test situation, became a priority. External expectations from schools and especially the child psychiatric services, were also mentioned.

Additionally, 'Some of it is also legitimacy, that you get legitimacy in your assessment [through including test results]'. It is reasonable to assume, as many others have noted, that the historical rationale of IQ tests as a technique to produce knowledge with an "objective quality", and as a way to assure the neutrality and "quality" of an assessment procedure, continue to underlie test use today.<sup>25</sup>

For Helga, the use of a test of general ability easily 'gives you a false feeling of safety', or certainty, regarding the results and analysis. Pointing back to her earlier statements about how assessment situations are more complex than what is found in the manual, she noted how many of the

---

25 Milowsky 1989; Birkemo 2009; Tveitnes 2018.

pupils that come in have a minority background. In her experience, WISC is used as an initial screening test in assessments in up to two-thirds of cases. This causes challenges; while one might be ‘conscious’ of the problem of assessing minority-language children, and ‘perhaps a bit more careful in the analysis, and including the typical phrasing in the report that the results *bør tolkes med forsiktighet* [should be interpreted carefully]’, one still conducts tests. In her opinion, one generally ‘does not test differently’.

Related to the rationale for tests as underpinning the legitimacy of assessments, it can be noted that the use of testing was put forward as a positive by EPS advisors across the board in cases and situations where the test results could be used to corroborate issues and problems observed either through observations or secondary sources—or in cases where a WISC indicated a normal test score and could be used to forcefully state that the problem lay elsewhere (in schools, in the classroom, in social adaptation, emotional difficulties). Separating and identifying the root causes of the problem are key tasks in an assessment, and it is for such purposes that tests like the WISC are thought to be useful, but only as “one piece of the puzzle”. This is similar to how they functioned in the early days of the first school psychological counseling offices.

When considering the statements on testing that introduce this section, from people working in or around Norway’s largest urban center, it is easy to imagine the use of WISC as an omnipotent, naturalized part of a normal assessment procedure. But as the EPS is organized differently, and educational backgrounds and test competence are also differently distributed regionally, this cannot be taken as a representative view of the EPS in general. Even within this same large, urban, regional office, there are likely several EPS advisors that would have sought to moderate the picture sketched above.

It must be mentioned that one can readily identify more measured approaches and descriptions of testing among the EPS. In smaller or more rural offices, some report that they do not have any employees with certification for the use of WISC and that they do not use it; and in the few cases where an WISC is needed, they refer to the district psychologist for testing and await the results for inclusion in their expert assessment. As an example, we can travel to the north of Norway, where Katarina—a local advisor in her late fifties with twenty years of experience—described a situation where most employees do not



have certified test competence, and where testing with a WISC was more infrequent. Far from it being a tool that would be first in line to be used during an assessment, she described them as 'being careful with WISC', pointing to the history of the office, where they had never had a psychologist employed. She described herself and her colleagues as pedagogues who were generally very 'restrictive with testing', at least in the case of intelligence tests. Instead, she narrated assessment procedures that emphasized observation, and more specific tests of language and reading, like LOGOS.

That said, several of their employees, including herself, have been certified in WISC in the last ten to fifteen years, and she reports an increase in use that she attributes to the wishes of the schools themselves, or the child and youth psychiatric services. Her own use of WISC, she describes as a last resort. She estimates that she does around five such tests a year in cases where other mapping tools and standard observational techniques fail to illuminate the problem, or in secondary instances, with re-referrals where existing special pedagogical measures fail to take effect. The restriction is a conscious one, in her words: 'A low score can equally often say something about the learning environment or other challenges... as a service we should be more systemically focused... but testing is individual, locating the problem in the child'.

As a smaller regional office, in contrast to larger urban centers, she described being closer to the child, knowing the children better, and having more time to observe classroom interactions. The tests have remained the same as they traveled, and even the understanding of these tests and the critique of testing are recognizable across place and time in the EPS. Still, not surprisingly, practices differ; it is, after all, people who test and are tested (here in the sense of individual subjects that always exist in, shape, and are shaped by larger socio-relational environments), despite the best efforts to circumvent this existential fact by recourse to statistics and standardized administrative procedures.

## Test as Professional Marker

Formalized assessment and test competence is usually found at the top of the qualification list of job advertisements from the EPS. In the words of one special pedagogy student specializing in specific learning

difficulties: 'I was in touch with the leader of [a local EPS office in a centrally located and populated municipality] to inquire into how to best qualify for a job in the EPS, what they were interested in was test competence; whether one is specialized in speech theory or developmental disability seemed less important'.

For EPS employees, the use of this type of standardized testing is a central aspect both in terms of the expectations and requirements of the work that they do, and it forms a key part of their professional identity:

[Testing] is kind of part of being a pedagogic-psychological advisor. A professional marker. Walking with this WISC suitcase over your shoulders, you know, that bag... When I was test-certified, then there was prestige in that. We were told that it is usually only certified psychologists that can test. That it was a privilege, and a proof [of] the disciplinary and scientific rigour of special needs pedagogy that we were allowed to test... It was [a] type of watershed moment when you get to the test-certification. Afterwards you become more professional, sort of... and you feel you are doing something important. —Helga

As frequently emphasized in test training across different courses, while a test alone is not enough of a foundation for a complete assessment—and they need to be used together with other tests and mapping tools, observations, and information gathered through conversations—the results of these tests, as seemingly objective measures of a students' general performance and ability independent from performance in school, do play a central role in many expert assessment procedures. This is the case historically and it is not limited only to Norway.

Carl Milofsky, who ethnographically traced the work of school psychologists in the greater Chicago area more than forty years ago, points to this, noting that it is not least in their capacity to 'impress both teachers and parents with their objectivity, accuracy, and authoritativeness' that intelligence tests have become powerful in assessment procedures and reports.<sup>26</sup> Testing, he argues, is the main identifiable signal of school psychological expertise, where it carries an air of 'precision and impersonality that makes them an object of respect by other educators'.<sup>27</sup> Tveitnes, when analyzing recent expert assessment

26 Carl Milofsky, *Testers and Testing: The Sociology of School Psychology* (New Brunswick: Rutgers University Press, 1989), p. 16.

27 Milofsky, *Testers*, p. 70.

reports from the EPS in Norway, points to the same dynamic at play. She argues for the continued existence of a knowledge hierarchy in expert assessment reports, where the ‘hard data’ of test results is given more weight than the ‘softer’ types of information gathered through observation and secondary accounts from teachers and parents.<sup>28</sup>

### Tests as Attempts at Objectivity: From Experience to Numbers that Communicate



Fig. 7.1 David Wechsler (1896–1981) not only created highly influential IQ tests, he also experimented with other physiological and psychological measurements. The identity of the exact test being administered in this photograph is unknown. Photographer unknown, 1955/New York University, School of Medicine, CC BY 4.0.

Psychology is a discipline that historically has readily offered its methods for use by others, be it governments, medicine, the military, and others.<sup>29</sup> In an extension of this, Emily Martin, for instance, recently showed how the methods of experimental psychology and adjoining assumptions—such as how we make choices as individuals, amplified by access to digital data, algorithms, and AI technology—are deeply embedded in our contemporary digital lives. ‘Old wine in new bottles’, according to Martin, as she notes how the power and influence of such methods rest on the strengths of past practices and technologies; ‘the

<sup>28</sup> Tveitnes, 2018

<sup>29</sup> Nicholas Rose, *Inventing Ourselves: Psychology, Power, and Personhood* (Cambridge: Cambridge University Press, 1998), pp. 87.

experiment, normative scales, training, and the extraction of data from subjects' that is part of constructing and reducing the world to a place 'occupied only by individual actions'.<sup>30</sup>

A key characteristic of such methods, at least in the tradition of psychometrics and experimental psychology, has been the ability to produce seemingly objective universal results from subjective experience in controlled settings. IQ tests, in the sense of being material and practical technological artifacts created through experimentation, standardized individual testing, and statistical analysis, and transported and adapted to fit similar objectives in new places and domains, fits the description well. Norway's EPS is only one such place where these technologies have traveled historically, and as an integrated part of the educational system, it is perhaps one of the less surprising places where they continue to be used today.

When one looks at the role and use of IQ tests like WISC in the EPS today, it is tempting to argue that it is here that their continued 'enchantment' as technological tools, in Alfred Gell's sense, still lies—in their capacity to sort unruly subjective experience and abilities into ordered and compatible scores and indexes for comparison in an assessment procedure.<sup>31</sup> As standardized tests, they embody the lure of the immediate and unambiguous characteristic of medical technology in general.<sup>32</sup> Part of their strength lies in concealment and disciplinary control. Their status as specialized psychometric scientific tools adds to the 'halo' effect of technical difficulty that Gell originally discussed in relation to aesthetic appreciation of art objects, Trobriand art, and garden magic. The complex technical mastery that lies behind scientific technologies like IQ tests is central to their efficacy. When at a loss to explain how an object or measure has come into the world, it is difficult to question or it easily becomes unquestionable.

IQ, as with any measurement, can function as a social fact, in the Durkheimian sense—capable of subjective coercion and control, in the sense that an IQ score "becomes" one's intelligence. As an added benefit,

---

30 Emily Martin, *Experiments of the Mind. From the Cognitive Psychology Lab to the World of Facebook and Twitter* (Princeton: Princeton University Press, 2021), p. 244.

31 Alfred Gell, *The Art of Anthropology: Essays and Diagrams* (London: Berg, 2006), especially pp. 159–186.

32 Eric J. Cassell, 'The Sorcerer's Broom: Medicine's Rampant Technology', *The Hastings Center Report*, 23 (6) (1993).

they result in numbers and categories, information that communicates well outwards, to schools, parents, and the surrounding welfare system. Numbers and statistics, and the social structures that produce them, after all, make the world legible for state control, manipulation, and governance.<sup>33</sup> This social facilitation of “trust in numbers”, it has been argued, is a defining feature of our society today. There is a fetishism in this drive to reduce qualitative information and experience to statistics and numbers, and the belief in this information as valid measurements and indicators capable of calculating, quantifying, and comparing such qualitative phenomena.<sup>34</sup> This quality is an important part of an IQ score. A number is easy to intuitively “understand”. As a numerical indicator it relies on institutional practices where ‘embedded theories, decisions about measures, and interpretations of the data are replaced by the certainty and lack of ambiguity of a number’.<sup>35</sup>

Such abstractions are always undermined by questions of morality and sociality, but it is the results that remain. To paraphrase Marilyn Strathern, who previously made this point in relation to her research on audit cultures: ‘An aura has come to surround numbers and, despite the caveats of professional [IQ testers], it is those unfamiliar with [IQ testing] who tend to sanctify them’.<sup>36</sup> Tests, and test results, are bound to be re-interpreted and come to possess a life of their own, and to be used for different justifications. It is no wonder, despite criticism, that these tests travel well. Their contemporary institutionalization and scale of use testify to their usefulness as administrative and diagnostic assessment tools regardless of how one interprets and uses the test scores, or whether the use is diagnostic or pedagogic.

Based on initial impressions, it is tempting to argue that a test like WISC and its results, while on the surface remaining the same, will be used and interpreted differently in the children and youth psychiatric services and the EPS, for instance. One can also expect to see a

33 James C. Scott, *Seeing Like a State* (Yale University Press, 1998).

34 Theodore Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Life* (Princeton: Princeton University Press, 1995); Chris Shore and Susan Wright, *Audit Culture: How Indicators and Rankings are Reshaping the World* (London: Pluto Press, 2024)

35 Sally Engle Merry, ‘Measuring the World: Indicators, Human Rights, and Global Governance’, *Current Anthropology*, 52 (3) (2011), 86, <https://doi.org/10.1086/657241>

36 Marilyn Strathern, *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy* (London: Routledge, 2000), pp. 8.

corresponding regional (and personal) and disciplinary differentiation within the services, as has been indicated above. Nevertheless, the test administration, and the resulting scores, remain (or should remain) the same. A lack of standardization across regional offices when it comes to test practice is reported as an internal problem and an issue of frequent discussion by counselors in the EPS. Despite this, there has so far been very little research into such differences, and a review of recent research on the EPS in Norway report surprisingly little research into how counselors deal with this important task given its social implications.<sup>37</sup>

This also means that the surrounding disciplinary expertise and knowledge that influence the use of tests, and the contexts and purposes for which they are employed must be kept in mind. But whether there are concrete differences, what they consist of, and to what degree testing with WISC by a special needs pedagogue in the EPS is different to that of a psychologist in the Children and Youth Psychiatric Service, for instance, remain elusive to us. It should also be noted that testing and the use of WISC-V in the EPS, for instance, includes a range of other testing technologies and mapping tools, together with observational procedures. In this larger nexus, the results of a test, and the weight given to them, will play out differently, depending not only on the background of the test administrator, but also on the case at hand and the test situation. A one-size-fits-all standardized technology of measurement, then, becomes a difficult thing to balance as an ideal and as practice. Exactly how this plays out necessitates further sustained research that follows expert assessment practice more closely as it is carried out, to delve deeper than the initial impressions offered here.

## Bibliography

- Aagaard, K. E., *Den språklige faktor: Pedagogisk-psykologisk utredning av barn med minoritetsspråklig bakgrunn* (PhD thesis, University of Oslo, 2011).
- Andrews, T. M. and B. Hustad, 'Bemanning I PP-Tjenesten: Tilstrekkelig for å oppfylle tjenestens mandat?', *Nordlandsforskning* (NF rapport nr.: 12/2022).
- Birkemo, A., *Pedagogisk-psykologisk rådgivning* (Unipub. Universitetet i Oslo, 2009).

---

<sup>37</sup> Torill Moen et al, 'The Norwegian Educational Psychological Service', *Nordic Studies in Education*, 38.2 (2018), 101–17.

- Bråten, I. and A. K. Henriksen, *Kvalifikasjonsproblem i tilknytning til den pedagogisk-psykologiske rådgivningstjenesten: Om utdanning av pedagogisk-psykologiske rådgivere* (Oslo: Seksjon for skolepsykologi, 1983).
- Cassell, E. J., 'The Sorcerer's Broom: Medicine's Rampant Technology', *The Hastings Center Report*, 23(6) (1993).
- Derksen, Maarten, 'Discipline, subjectivity and personality: an analysis of the manuals of four psychological tests', *History of the Human Sciences*, 14 (1) (2001), <https://doi.org/10.1177/095269510101400102>
- Fjeld, F. E., *Bruk av WISC-IV i pedagogisk psykologisk tjeneste: Hvordan vurderer ansatte i PPT WISC-IV som et verktøy i utredning av barn?* (MA thesis, Universitetet i Oslo, 2015).
- Gell, A., *The Art of Anthropology: Essays and Diagrams* (London: Berg, 2006).
- Helmikstøl, Ø. and N. Skjerdingsstad, 'Testfelt uten styring', *Tidsskrift for Norsk psykologforening*, 58.1 (2021).
- Hesselberg, F. and S. von Tetzchner, *Pedagogisk-psykologisk arbeid* (Gyldendal, 2016).
- Martin, E., *Experiments of the Mind. From the Cognitive Psychology Lab to the World of Facebook and Twitter* (Princeton: Princeton University Press, 2021), p. 244, <https://doi.org/10.23943/princeton/9780691230719.002.0003>
- Merry, S. E., 'Measuring the World: Indicators, Human Rights, and Global Governance', *Current Anthropology*, 52(3) (2011), <https://doi.org/10.1086/657241>
- Meyer, G. J. et al, 'Psychological Testing and Psychological Assessment: A Review of Evidence and Issues', *American Psychologist*, 56.2, <https://doi.org/10.1037/0003-066X.56.2.128>
- Milofsky, C., *Testers and Testing: The Sociology of School Psychology* (New Brunswick: Rutgers University Press, 1989).
- Moen, T. et al, 'The Norwegian Educational Psychological Service', *Nordic Studies in Education*, 38.2 (2018), <https://doi.org/10.18261/issn.1891-5949-2018-02-02>
- Nordahl, T. et al, *Inkluderende Fellesskap for barn og unge: Ekspertgruppen for barn og unge med behov for særskilt tilrettelegging* (Fagbokforlaget, 2018).
- Phil, J., *Etnisk mangfold i skolen: Det sakkyndige blikket* (Universitetsforlaget, 2<sup>nd</sup> ed., 2010).
- Porter, T., *Trust in Numbers: The Pursuit of Objectivity in Science and Life* (Princeton: Princeton University Press, 1995).
- Mülberger Rogele, A., 'Biographies of a Scientific Subject: The Intelligence Test', in *Oxford Research Encyclopedias: Psychology* (Oxford: Oxford University Press, 2020), <https://doi.org/10.1093/acrefore/9780190236557.013.694>

- Rose, N., *Inventing Ourselves: Psychology, Power, and Personhood* (Cambridge: Cambridge University Press, 1998).
- Ryder, T. M., 'Testkvalitetsprosjektet - del 1: Norske psykologers testholdninger og testbruk', *Tidsskrift for Norsk psykologforening*, 58.1 (2021).
- 'Testkvalitetsprosjektet—del 2: Tester med behov for kvalitetstiltak', *Tidsskrift for Norsk psykologforening*, 58.2 (2021).
- Scott, J. C., *Seeing Like a State* (Yale University Press, 1998).
- Shore, C. and S. Wright, *Audit Culture: How Indicators and Rankings are Reshaping the World* (London: Pluto Press, 2024).
- Strathern, M., *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy* (London: Routledge, 2000), <https://doi.org/10.4324/9780203449721>
- Tveitnes, M. S., *Sakkyndighet med mål og mening: En analyse av sakkyndighetens institusjonaliserte kjennetegn; et grunnlag for refleksjon og endring* (PhD thesis, University of Stavanger, 2018).
- Wechsler, D., *The Measurement of Adult Intelligence* (Williams & Wilkins Co., 1939).





## 8. ‘Children Got Slightly Smarter with Fish for Lunch’: The WPPSI Test, Randomized Trials, and Optimized Kindergartens

*Susanne Bauer*

---

This chapter follows the Wechsler tests to a rather unexpected place—a nutrition trial in Norwegian kindergartens. During the spring of 2015, more than 200 children in thirteen kindergartens in Bergen completed the Wechsler Preschool and Primary Scale of Intelligence, 3rd edition (WPPSI-III). This endeavor of testing for cognitive capacities was part of an intervention trial during which the children ate standardized fatty fish or meat lunches for six weeks. The results of the trial were published three years later and local newspapers and research websites covered the findings, boldly claiming that ‘lots of fat fish makes children think faster’.<sup>1</sup> The focus on fish in the trial and the reporting to the wider public points us to the core of Norway’s bioeconomy aspirations.<sup>2</sup> Fish and aquaculture, especially farmed salmon, is promoted not only as healthy, but also as a resource poised to constitute the new post-fossil bioeconomy—the vision for Norway’s transition away from oil and gas.<sup>3</sup>

---

1 ‘Mye fet fisk får barn til å tenke fortere. Det sies at omega 3 er bra for hjernen. Når har bergensforskere funnet at mye fet fisk får barn å tenke fortere’, *Stavanger Aftenblad*, 12 March 2018; ‘Barn ble litt smartere med fisk til lunsj’, Norwegian online science magazine forskning.no, 12 March 2018 (<https://forskning.no/partner-barn-og-ungdom-havforskningsinstituttet/barn-ble-litt-smartere-med-fisk-til-lunsj/287470>).

2 See also for instance: ‘Nå har bergensforskere funnet at mye fet fisk får barn å tenke fortere’, *Stavanger Aftenblad*, 12 March .2018.

3 Markus Bugge, Teis Hansen, and Antje Klitkou, ‘What Is the Bioeconomy?’, in *From Waste to Value: Valorisation Pathways for Organic Waste Streams in Circular Bioeconomies*, ed. by Antje Klitkou, Arne Martin Fevolden, and Marco Capasso (Routledge, 2019), pp. 19–50; Kristin Asdal, and Tone Huse, *Nature-Made Economy: Cod, Capital and the*

Seeking to understand how kindergarten lunches became research settings for enhancing intelligence through fish meals, this chapter will examine these attempts at societal problem-solving and (post-)welfare health governance.<sup>4</sup> In doing so, it follows one version of the Wechsler tests, the WPPSI-III, and examines how it became part of reconfiguring the relations between eating, welfare, inclusion, and cognition. To analyze such testing practices, I take inspiration from studies of biopolitics, care, and (bio)medicalization.<sup>5</sup> The rise of quantification and measurement in the sciences and psychology has led to scientific understandings of the normal and the pathological being established, contested, negotiated—and they have often persisted. Medicalization has expanded beyond medicine and encompasses psychological sciences and pedagogy, including knowledge about intellectual disability. Along with the health sciences, these fields have privileged particular formats of scientific expertise, organized around standardized devices for testing.<sup>6</sup> With the standardization of intelligence testing, transnational organizations of applied psychology contributed to building what emerged as global ‘testing cultures’ in the twentieth century.<sup>7</sup>

---

*Great Economization of the Ocean* (MIT Press, 2023).

- 4 Signild Vallgård, ‘Governing People’s Lives: Strategies for Improving the Health of the Nations in England, Denmark, Norway and Sweden’, *European Journal of Public Health*, 11.4 (2001), 386–92, <https://doi.org/10.1093/eurpub/11.4.386>; Karsten Vrangbæk, ‘The Regulation of Health Care in Scandinavia’, in *Professional Health Regulation in the Public Interest*, ed. by John Martyn Chamberlain, Mike Dent, and Mike Saks, International Perspectives, 1st edn. (Bristol University Press, 2018), pp. 61–76, <https://doi.org/10.2307/j.ctv1xxs5q.10>; M. Buchardt, P. Markkola, and H. Valtonen, ‘Introduction: Education and the Making of the Nordic Welfare States’, in *Education, State and Citizenship* (University of Helsinki, 2013), pp. 7–30.
- 5 Michel Foucault, *The Birth of Biopolitics: Lectures at the Collège de France, 1978–1979*, trans. by G. Burchell. (New York: Palgrave Macmillan, 2008); *Care in Practice*, ed. by Annemarie Mol, Ingunn Moser, and Jeannette Pols; *On Tinkering in Clinics, Homes and Farms* (transcript Verlag, 2010), <https://doi.org/10.1515/transcript.9783839414477>; Vincanne Adams, Michelle Murphy, and Adele E Clarke, ‘Anticipation: Technoscience, Life, Affect, Temporality’, *Subjectivity*, 28.1 (2009), 246–65, <https://doi.org/10.1057/sub.2009.18>.
- 6 Daniel Tröhler, ‘The Medicalization of Current Educational Research and Its Effects on Education Policy and School Reforms’, *Discourse: Studies in the Cultural Politics of Education*, 36.5 (2015), 749–64, <https://doi.org/10.1080/01596306.2014.942957>; Bjørn Hamre, Thom Axelsson, and Kari Ludvigsen, ‘Psychiatry in the Sorting of Schoolchildren in Scandinavia 1920–1950: IQ Testing, Child Guidance Clinics, and Hospitalisation’, *Paedagogica Historica*, 55.3 (2019), 391–415, <https://doi.org/10.1080/00309230.2019.1597137>
- 7 John Carson, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750–1940* (Princeton University Press, 2006); Christian Ydesen, ‘Crafting the English Welfare State—Interventions by Birmingham Local Education

To examine the lingering categorizations mobilized in these tests, I draw on literature from the history of statistics, and science and technology studies (STS).<sup>8</sup> I attend to how testing practices are made routine and interrogate the widespread appearance of ableist projects under the label of optimization.<sup>9</sup> Foregrounding the heterogeneous modes of testing and trials in educational settings also speaks to histories of 'state crafting':<sup>10</sup> While, from its inception, the welfare state was invested in the optimization of people's health to keep the workforce intact, the modern social sciences reflected and participated in pertinent debates.

This chapter zooms in on the material practices of cognitive testing within intervention trials in nutrition research. It describes how the study designs and scientific practices reify certain arrangements and categorizations in the tests themselves, their settings, and their modes of administration. I will show how the test is used as a device that can be plugged into the battery of other inscription techniques, stabilized by and stabilizing the data generation, tabulation, and statistical analysis. Situating itself within STS, this contribution follows the WPPSI tests in action and brings to the fore the logistics and specific 'boxing practices'<sup>11</sup> that are part of the ordering work entailed in administering Wechsler tests in the setting of nutrition trials in Norwegian kindergartens.

The following sections revolve around the test itself, and the orderings enacted through testing, the work of study design, the published tables displaying the outputs of statistical analysis, and the reassembling of the testing performs. I analyze tests as calculative devices, as well as the statistical tables as 'numerical boxes' that enact specific data orderings. Thus, this chapter contributes to understanding the role of IQ testing in post-welfare-state science policies through an analysis of the micro-practices of the WPPSI-III test in a nutrition trial in Norwegian kindergartens.

---

Authorities, 1948–1963', *British Educational Research Journal*, 42.4 (2016), 614–30.

8 Steven M. Stigler, *The History of Statistics: The Measurement of Uncertainty before 1900* (Belknap Press of Harvard University Press, 1986); Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (1995); Alain Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning* (Harvard University Press, 1998); Carson.

9 Fiona Kumari Campbell, *Contours of Ableism: The Production of Disability and Ableness* (Palgrave Macmillan, 2009); Ydesen.

10 Ydesen.

11 Susanne Bauer, Martina Schlünder, and Maria Rentetzi, *Boxes. A Field Guide* (Mattering Press, 2020).

‘Things You Put Things “In”’: Test Puzzles and Fish Trials

The WPPSI test, like the other Wechsler tests, literally comes in boxes with test kits, materials, and manuals (see Fig. 8.1). Among its modules, the WPPSI includes a task sheet titled ‘Containers. Things to put things in’.<sup>12</sup> This is just one of many tasks and materials for testing children’s ability to grasp object sets and orderings. ‘Putting things into containers’—in the broadest sense of the term—plays an important role in the material practices of the test itself, but also in any categorization and ordering exercise.<sup>13</sup> In a more general sense, putting things into boxes is an epistemic practice that forms part of many analytic operations. These have been universalized in Euro-American sciences and incorporated into educational testing kits.<sup>14</sup>



Fig. 8.1 WPPSI-III Test box; WPPSI-III Manual; example from ‘Concept Grouping’ part of the performance IQ test (picture concepts, supplemental sections). Photographs by the author. CC BY-NC.

12 WPPSI test, English version.  
13 Bauer, Schlünder, and Rentetzi.  
14 Geoffrey C. Bowker and Susan Leigh Star, ‘Sorting Things out: Classification and its Consequences’, *Inside Technology* (MIT Press, 1999).

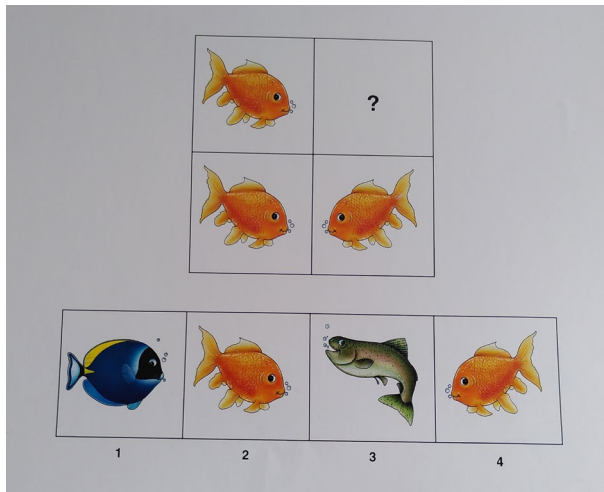


Fig. 8.2 Test examples: Fish puzzles in WPPSI-III test (materials from the Swedish and Norwegian version). Photographs by the author. CC BY-NC.

The WPPSI test set contains items specific to age groups; these illustrations represent the four-to-seven-year age range. The set comprises several modules (such as vocabulary, word reasoning, or block design), with visual materials and objects for conducting the test (see Figs. 8.1 and 8.2). The test calculates a general score as well as verbal, performance, processing, and general language subscores. As the most established and validated standard method, the WPPSI test set also made its way into a series of Fish Intervention Studies (FINS)

conducted in schools and kindergartens (FINS-KIDS) in the 2010s, when a multidisciplinary team of scientists conducted nutrition trials in kindergartens in Bergen.<sup>15</sup> At the time of the work for the FINS-KIDS study in spring 2015, the WPPSI-III version, later replaced by WPPSI-IV,<sup>16</sup> was the state-of-the-art test.<sup>17</sup> For the FINS-KIDS nutrition trial, researchers included all fourteen different modules and additionally used a nine-hole peg test for fine motor coordination, measuring the speed with which the tested child rearranged certain items.

Conducting such comprehensive tests is costly due to the complex logistics and the required meticulous preparation, from recruitment and randomized allocation of meals to the practicalities of test administration. For the FINS-KIDS project, this was organized through conducting the tests for sixty to ninety minutes in 'separate rooms in the kindergartens, with noise kept to a minimum'.<sup>18</sup> Test administrators were unaware of who was in either group. Each aspect needed to be managed—a caterer for lunch delivery; research assistants to track who got which lunches, ensure that no food was shared, and document compliance; and trained WPPSI test administrators and biomedical practitioners to oversee the test and draw blood samples. To additionally validate results, researchers had also included biomarker studies, so they collected—from those children whose parents had additionally agreed to this part—blood, urine, and hair samples.

The two different lunches the caterer prepared and delivered for the study comprised one set of meals containing 50–80 grams of fatty fish such as mackerel, herring, or salmon; the second group was given meals with an identical weight of meat instead of fish.<sup>19</sup> The intervention lasted for six weeks. As the goal was to quantify whether, and to what

---

15 Jannike Øyen et al, 'Fatty Fish Intake and Cognitive Function: FINS-KIDS, a Randomized Controlled Trial in Preschool Children', *BMC Medicine*, 16.1 (2018), p. 41, <https://doi.org/10.1186/s12916-018-1020-z>

16 See Caspersen, Kyllingstad, this volume's introduction.

17 Wechsler tests were developed in the US in 1967 and, as standardized sets for psychometric testing, they gained ground in education and mental health diagnostics, and were widely adopted and adapted to different languages. Some of the adaptations to Norwegian settings became controversial; at the time of conducting the study (in 2015), the test version was routinely used in education research.

18 Øyen et al, 2018, p. 3.

19 Øyen et al, 2018.

extent, a change to more fish would improve the children's cognitive function, children completed the WPPSI twice—at baseline and after the intervention.

The project of creating standardized test conditions is necessarily in tension with a kindergarten setting, making the children's shared everyday procedures a potential disturbance to the trial. To make the study happen, a broad range of different people—educators, kindergarten teachers, parents, caterers, psychologists and research assistants—needed to be on board and work according to the study's protocol. For the tests to happen at all, their varied processes and labor needed to be synchronized. Researchers prepared the logistics such that they could administer the WPPSI tests in separate rooms of the kindergarten. Having children eat a standardized lunch can prove even more difficult than motivating them to take a test. Scientists report matter-of-factly, as expected for a scientific paper, a range of 'compliance issues', such as 'dislike of the food', which made some of the results inconclusive. This makes visible how an "ideal" study design that demands a controlled experimental setting is in tension with a kindergarten setting with children being unruly research subjects no matter if parents have agreed to their participation in the study.

Next, I take up the test's "fish puzzles" and the feature of "putting things into boxes" in order to shift the analysis to the performativity implicated in the design and execution of this intervention trial. In scientific knowledge, production design processes bring about and enact particular onto-epistemic orderings with which researchers then continue to work.<sup>20</sup> Here, the notion of putting things into containers is taken further to analyze testing and statistical practices, the enrolment of test subjects, as well as the categorizations that tests and testing enact.

---

20 Helen Verran, 'Re-Imagining Land Ownership in Australia', *Postcolonial Studies*, 1.2 (1998), 237–54, <https://doi.org/10.1080/13688799890165>; Helen Verran, 'Decomposing Numbers', *HAU Journal of Ethnographic Theory*, 8.1–2 (2018), 23–26.



## Study Designs: Enrolling Kindergartens into a Nutrition Trial

Nutrition recommendations emphasize the health benefits of omega-3 fatty acids, a feature that has been at the core of health advice for some time.<sup>21</sup> To consumers, farmed salmon has often been advertised both as high-protein fish and, with its content of omega-3 fatty acids, as beneficial to cognitive development and enhancers of human performance. A public report in 2022 summarized the few existing studies on children's cognitive performance, including the FINS studies.<sup>22</sup> Most publications emerging from the FINS-KIDS study had eight to fifteen co-authors from a range of different disciplines, backgrounds and positions.

The study design comprised the standardized set-up for an experimental situation to represent and analyze a multifactorial process. To achieve this multidisciplinary endeavor, FINS brought together researchers from different disciplines and institutions. The various institutions involved in the study ranged from pedagogical research institutes to seafood research, and both public and private organizations—but it was the fish industry that had initiated and funded large parts of the study. The design of the study involved experts from seafood research, nutrition science, and child development, to neuropsychology, environmental chemistry and toxicology, each with different disciplinary traditions, viewpoints, and concerns.

The planning phase and discussions about study design were when the interdisciplinary research consortium made the key decisions that would later guide all aspects and logistics. This comprised bringing together different disciplinary components; deciding on research questions, main hypotheses, and variables; and planning analytic strategies. The set-up needed to adhere to the scientifically sound methods of evidence-based medicine and evidence-based policymaking: the clinical intervention trial.<sup>23</sup> The intervention trial focused on investigating how unsaturated

---

21 In their overall assessments, public health institutions recommend eating fish as a healthy protein source, even though fatty fish accumulate toxic persistent organic pollutants present in the water, which by this pathway enter the human body (Bustad et al, 2008).

22 VKM Vitenskapskomiteen for Mat og Miljø, 'Benefit and risk assessment of fish in the Norwegian diet—Scientific Opinion of the Steering Committee of the Norwegian Scientific Committee for Food and Environment' (2022), <https://nmbu.brage.unit.no/nmbu-xmlui/handle/11250/3060289?show=full&locale-attribute=en>

23 D. Sackett et al, *Evidence Based Medicine* (Churchill Livingstone, 1995); Barrie Margetts et al, 'Developing an Evidence-Based Approach to Public Health Nutrition:

fatty acids in fish meals might improve cognitive function when compared to meat meals. Its study design—the Randomized Controlled Trial—has been widely recognized as the gold standard for producing rigorous evidence and testing hypotheses from clinical research to public health. The randomized trial requires specific features, such as randomization, different study arms, and “blinded” researchers administering the test without knowing which food they had. The investigators responsible for the study designs translated those formal requirements and respective disciplinary methods into study protocols and logistics.

The study design phase is the (only) step when principal investigators and researchers from different disciplines discuss content and research questions, problematize methods and metrics, arrange their different components, divide responsibilities, negotiate over diagrams and methods, and progress towards operationalizing this expertise for the study. Once decided upon and integrated, these decisions cannot be modified while the study is being carried out.

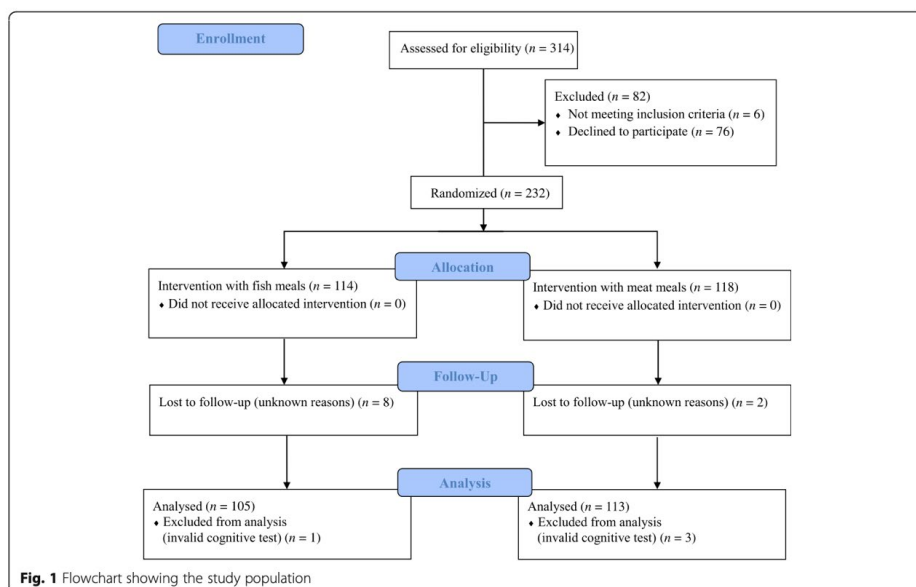


Fig. 8.3 Design box. Organizing the intervention and the study population (plugging Wechsler tests into a nutrition trial). Published in Jannike Øyen, Ingerid Kvestad, Lisa Kolden Midtbø, et al. 'Fatty Fish Intake and Cognitive Function: FINS-KIDS, a Randomized Controlled Trial in Preschool Children', *BMC Medicine*, 16.1 (2018). CC BY 4.0.

Figure 8.3, reproduced from the journal publication, presents the “sample”—in other words, the “study population”—of kindergarten kids. The group is divided into two study arms, with the boxes arranged in a way that presents the information as relevant for the statistical data processing and the interpretation of risk estimates. Constructing, implementing, and aligning such a study with the real life of educational institutions requires considerable resources. Aligning and carrying out the study in everyday life and settings like kindergartens also requires significant planning, consideration, personnel, and time. Once the study design was in place, pilot studies were done to test the questionnaires and develop the logistics. The recruitment phase is a major part of the preparation, especially for vulnerable study populations and at a place like a kindergarten, and it included obtaining individual consent from kindergartens and convincing parents to allow their children to participate in the study. The recruitment plan actively constructs the study population—sometimes researchers already address known or expected asymmetries, for instance by contacting more members of a lower socioeconomic background who are less likely to have participated in health studies.

The study population boxes are assigned and completed in accordance with the Randomized Clinical Trial and its analysis plan. Still, within these neat boxes, some troubles and disturbances—“non-compliance”—can be traced, for instance when participants are “lost to follow-up”. This is quantified and listed in order to assess any systematic error (bias), and is important to assure the quality of the study. Too many losses to follow-up in terms of dropouts or disturbances would jeopardize the validity of the study. In our case of the kindergarten fish trial, “lost to follow-up” can mean that there is only one measurement for the WPPSI test—because a participant dropped out or one of the measurements was not gathered according to protocol. Importantly, the research subjects may not be as docile as imagined by the planners, and their way of behaving in relation to the test may not fall in line with what the study prescribed. Some children may simply prefer other food than the assigned standard lunches, or they may want to leave the test setting and go about their own play after a while. Timing and test duration are important factors to consider in planning the tests.

Carrying out such intervention trials has worked to realign educational systems with evidence-based public health and policymaking—and also with the role of aquaculture in Norway's bioeconomy aspirations. This set of nutrition trials, with funding from both the aquaculture industry and the Norwegian Research Council, are one example of how science policy and research funding is laid out to promote collaboration between the private and public sector in Norway. The institutional set-up included the *Nasjonalt institutt for ernærings- og sjømatforskning* [NIFES, the National Institute of Nutrition and Seafood Research], which in 2018 merged into *Hauforskningsinstituttet* [Institute of Marine Research], affiliated with the Ministry of Trade, Industry, and Fisheries. In published papers, the consortium addressed the question of marketing interests, given the funding sources listed which included the Norwegian Seafood Research Fund, Marine Harvest A/S, Lerøy A/S, and Pelagia A/S. Complying with journal requirements, ethics protocols, and trial registration, the authors disclosed that the involved companies provided the mackerel and herring for the trial free of charge, but stated that they 'had no role in the design of the study, in the collection, analyses, or interpretation of data, in the manuscript writing, or in the decision to publish the results.'<sup>24</sup> In view of potential industry influence on these scientific studies, the researchers stressed their efforts to balance any conflicts of interest by insisting on strong adherence to a strict and rigorous scientific methodology.

Such enactments of the tests participate in experimentalizing education, nutrition, and cognition at the same time. With intervention trials, modeled on the Randomized Clinical Trial format, the main function of including the WPPSI test is to generate reliable and comparable data, using the standard in the field to produce hard evidence. Yet, independently of their results, practicing scientific trials as experiments within schools and kindergartens have an impact beyond testing whether fish consumption enhances cognitive abilities. They habituate professionals in educational institutions and citizens to specific modes of evidence-based governance. They are also a public exercise regarding the possibilities of optimizing performance and the means to do so in institutions like kindergartens and schools. The test

---

24 Øyen et al, p. 12.

gives credibility and is part of enabling and normalizing living within controlled experiments.

The interdisciplinary collaboration for the set-up of the trial is the moment of encounters between different disciplines, in which researchers bring together of different elements and modules into one study design. It is here that the psychometric measurement—and the Wechsler test—is plugged into a randomized intervention study. Yet, the Wechsler test is just one of possible tools for such a randomized intervention trial with the overall goal of studying whether fish meals indeed enhance cognitive function in children. Researchers responsible for psychometric testing stress that the Wechsler test is a well-studied, researched, validated test system translated into Norwegian. It is considered a tool that, depending on the research question, can be implemented and interpreted reliably, and is hence the best choice, as not many test batteries are readily available in Norwegian. Any additional translations and validations of other test systems needed would lead to further costs, making the already expensive tests for the study unrealistic.

Once designed, the study protocol demands execution by the book—any change may introduce inconsistencies or bias and disable interpretation. Every step thus needs to be carried out according to protocol. Even with the reliable RCT design applied to this kindergarten setting, interpretation is still needed to deal with uncertainties. These uncertainties arise as much of the testing calibrations come from different places, such as the US, and different times. They thus need careful consideration when used in new settings, such as twenty-first-century Norway. The next section turns to the resulting data and pertinent orderings through tables, as well as their translations and effects.

## Columns and Rows: Making Tables for Risk-Benefit Optimization

Producing, boxing, and analyzing data is at the core of scientific practices that extract numbers through measurement and put

numbers into columns and rows. Tables have a central position in scientific publications—health scientists use tables as descriptive lists of demographics of their study population. This information helps researchers interpret results, and assess potential biases and the generalizability of the findings. Result sections in scientific publications contain both textual summaries as well as tables reporting statistical analyses. Tables often contain epistemic framings, and columns and rows are containers that can both inform and result from statistical modeling. The columns and rows take up the boxes of the flow chart from the construction of the study population. Study designs group the variables into determinants and outcome relations in order to empirically test hypotheses and describe the relations between variables. Indeed, tables work as ordering devices par excellence.<sup>25</sup> Tables break down these boxes further to subgroups, with the procedures of data analysis and model adjustments for confounding and interacting variables. At times, researchers add more layers to the main hypothesis under study, through exploratory subgroup analyses. Scientists use the latter for hypothesis generation to be tested in new studies designed to specifically address these emergent hypotheses.

Building on and integrating various modules, from nutritional epidemiology to psychological testing, the different inscription devices make up an experimental setting that produces standardized data. Researchers use the resulting data to statistically test for associations between fish meals and cognitive performance, i.e., by examining for significant differences between the fish and the meat group. Figure 8.4 (a, b) shows two published tables, presenting statistical analyses of the data generated in the intervention trial. These tables are boxes that summarize results of statistical testing. The presentation is specific to the research questions and geared to test the respective associations between fish lunches (versus meat lunches) and 'change in the Wechsler Preschool Scale of Intelligence', as operationalized in the study design (Fig. 8.4a). Figure 8.4b displays the WPSSI post-intervention scores with adjustments for different covariates.

---

25 Bowker and Star; Bauer, Schlünder, and Rentetzi.

**Table 3** Predicted change in Wechsler Preschool and Primary Scale of Intelligence, 3rd edition (WPPSI-III) raw scores after study meals with fish (n = 105) and meat (n = 113)

| Intervention           | Main analyses |                |                | Models adjusted for         |        | Sub-analyses                            |          |                                                                           |          |
|------------------------|---------------|----------------|----------------|-----------------------------|--------|-----------------------------------------|----------|---------------------------------------------------------------------------|----------|
|                        | Pre Mean (SD) | Post Mean (SD) | p <sup>a</sup> | Pre-score, age <sup>b</sup> |        | Pre-score, age, compliance <sup>c</sup> |          | Pre-score, age, compliance, interaction treatment*compliance <sup>d</sup> |          |
|                        |               |                |                | Change Mean (95% CI)        | p      | Change Mean (95% CI)                    | p        | Change Mean (95% CI)                                                      | p        |
| PPVT-III               |               |                |                |                             |        |                                         |          |                                                                           |          |
| Total raw score        |               |                |                |                             |        |                                         |          |                                                                           |          |
| Fish                   | 145.3 (35.0)  | 162.7 (36.2)   | < 0.0001       | 17.7 (14.8 to 20.7)         | 0.97   | 20.4 (17.5 to 23.3)                     | 0.0060   | 21.9 (19.4 to 24.5)                                                       |          |
| Meat                   | 141.7 (33.5)  | 159.8 (33.1)   | < 0.0001       | 17.8 (15.0 to 20.6)         |        | 15.2 (12.4 to 18.0)                     |          | 17.2 (14.7 to 19.8)                                                       |          |
| Model fit <sup>e</sup> |               |                |                |                             |        |                                         | < 0.0001 |                                                                           | < 0.0001 |
| Verbal raw score       |               |                |                |                             |        |                                         |          |                                                                           |          |
| Fish                   | 56.6 (13.4)   | 60.4 (14.2)    | < 0.0001       | 3.8 (2.6 to 5.0)            | 0.59   | 4.7 (3.6 to 5.8)                        | 0.11     | 4.9 (3.8 to 6.1)                                                          |          |
| Meat                   | 55.9 (12.7)   | 60.2 (12.5)    | < 0.0001       | 4.3 (3.1 to 5.4)            |        | 3.4 (2.4 to 4.5)                        |          | 3.7 (2.6 to 4.8)                                                          |          |
| Model fit <sup>e</sup> |               |                |                |                             | 0.0061 |                                         | < 0.0001 |                                                                           | 0.072    |
| Information            |               |                |                |                             |        |                                         |          |                                                                           |          |
| Fish                   | 23.8 (3.2)    | 24.8 (3.0)     | < 0.0001       | 1.0 (0.6 to 1.4)            | 0.63   | 1.1 (0.7 to 1.4)                        | 0.90     | 1.1 (0.7 to 1.4)                                                          |          |
| Meat                   | 23.5 (3.2)    | 24.7 (3.0)     | < 0.0001       | 1.1 (0.8 to 1.5)            |        | 1.0 (0.7 to 1.4)                        |          | 1.0 (0.7 to 1.4)                                                          |          |
| Model fit <sup>e</sup> |               |                |                |                             |        |                                         | 0.067    |                                                                           | 0.95     |
| Vocabulary             |               |                |                |                             |        |                                         |          |                                                                           |          |
| Fish                   | 18.0 (6.5)    | 19.2 (6.9)     | 0.015          | 1.1 (0.3 to 1.9)            | 0.99   | 1.7 (0.9 to 2.4)                        | 0.0468   | 1.7 (1.0 to 2.5)                                                          |          |
| Meat                   | 18.4 (6.1)    | 19.5 (5.9)     | 0.006          | 1.1 (0.4 to 1.9)            |        | 6 (-0.1 to 1.3)                         |          | 0.7 (-0.1 to 1.4)                                                         |          |
| Model fit <sup>e</sup> |               |                |                |                             |        |                                         | < 0.0001 |                                                                           | 0.36     |
| Block Reasoning        |               |                |                |                             |        |                                         |          |                                                                           |          |
| Fish                   | 14.8 (5.2)    | 16.4 (5.5)     | < 0.0001       | 1.8 (1.1 to 2.4)            | 0.50   | 2.0 (1.4 to 2.7)                        | 0.54     | 2.2 (1.5 to 2.8)                                                          |          |
| Meat                   | 13.9 (4.9)    | 16.0 (5.1)     | < 0.0001       | 2.1 (1.4 to 2.7)            |        | 1.8 (1.1 to 2.4)                        |          | 1.9 (1.3 to 2.6)                                                          |          |
| Model fit <sup>e</sup> |               |                |                |                             |        |                                         | 0.0001   |                                                                           | 0.036    |
| Performance raw score  |               |                |                |                             |        |                                         |          |                                                                           |          |
| Fish                   | 51.7 (10.7)   | 57.3 (10.9)    | < 0.0001       | 6.0 (4.7 to 7.3)            | 0.65   | 6.4 (5.2 to 7.7)                        | 0.16     | 6.6 (5.4 to 7.9)                                                          |          |
| Meat                   | 49.2 (10.4)   | 55.2 (10.1)    | < 0.0001       | 5.6 (4.4 to 6.8)            |        | 5.2 (4.0 to 6.4)                        |          | 5.4 (4.2 to 6.6)                                                          |          |
| Model fit <sup>e</sup> |               |                |                |                             |        |                                         | 0.0018   |                                                                           | 0.13     |
| Block Design           |               |                |                |                             |        |                                         |          |                                                                           |          |
| Fish                   | 24.6 (3.6)    | 26.2 (3.4)     | < 0.0001       | 1.7 (1.3 to 2.1)            | 0.068  | 1.8 (1.3 to 2.2)                        | 0.0269   | 1.9 (1.4 to 2.3)                                                          |          |
| Meat                   | 24.3 (3.7)    | 25.5 (3.4)     | < 0.0001       | 1.1 (0.7 to 1.6)            |        | 1.1 (0.6 to 1.5)                        |          | 1.1 (0.7 to 1.6)                                                          |          |
| Model fit <sup>e</sup> |               |                |                |                             |        |                                         | 0.14     |                                                                           | 0.15     |

I. Kvestad et al. *Environment International* 121 (2018) 1098–1105

**Table 4** Post-intervention WPPSI-III raw scores after study meals with fish (n = 101) or meat (n = 109) adjusted for THHg and/or fish/meat consumption.

| Models <sup>a</sup>                           | N   | WPPSI-III post-scores (95% CI) |                      | p-Value <sup>b</sup> |
|-----------------------------------------------|-----|--------------------------------|----------------------|----------------------|
|                                               |     | Fish                           | Meat                 |                      |
| Total                                         |     |                                |                      |                      |
| Age                                           | 210 | 162.6 (156.5, 168.6)           | 160.0 (154.1, 165.9) | 0.475                |
| Age and THHg                                  | 210 | 160.4 (154.1, 166.7)           | 161.8 (155.7, 167.8) | 0.741                |
| Age and fish/meat consumption                 | 210 | 164.0 (160.7, 167.4)           | 158.3 (155.0, 161.5) | 0.003                |
| Age, THHg and fish/meat consumption           | 210 | 163.8 (160.3, 167.3)           | 158.5 (155.1, 161.9) | 0.017                |
| Age, THHg, fish/meat consumption, sex and SES | 186 | 164.5 (160.9, 168.1)           | 159.0 (155.6, 162.4) | 0.008                |
| Verbal                                        |     |                                |                      |                      |
| Age                                           | 210 | 60.2 (57.9, 62.5)              | 60.1 (57.8, 62.3)    | 0.914                |
| Age and THHg                                  | 210 | 59.4 (57.0, 61.9)              | 60.8 (58.5, 63.1)    | 0.453                |
| Age and fish/meat consumption                 | 210 | 61.2 (58.9, 63.5)              | 59.2 (56.9, 61.4)    | 0.223                |
| Age, THHg and fish/meat consumption           | 210 | 60.5 (58.1, 63.0)              | 59.8 (57.4, 62.1)    | 0.692                |
| Age, THHg, fish/meat consumption, sex and SES | 186 | 61.1 (58.8, 62.4)              | 59.9 (58.7, 61.1)    | 0.164                |
| Performance                                   |     |                                |                      |                      |
| Age                                           | 210 | 56.4 (54.9, 57.9)              | 56.4 (55.0, 57.8)    | 0.973                |
| Age and THHg                                  | 210 | 56.4 (54.9, 57.9)              | 56.4 (54.9, 57.8)    | 0.966                |
| Age and fish/meat consumption                 | 210 | 56.8 (55.4, 58.2)              | 55.9 (54.5, 57.3)    | 0.315                |
| Age, THHg and fish/meat consumption           | 210 | 57.5 (55.4, 59.6)              | 55.4 (53.3, 57.4)    | 0.097                |
| Age, THHg, fish/meat consumption, sex and SES | 186 | 56.7 (55.2, 58.2)              | 56.3 (54.9, 57.6)    | 0.659                |
| Processing speed                              |     |                                |                      |                      |
| Age                                           | 210 | 45.1 (42.8, 47.4)              | 44.3 (42.1, 46.5)    | 0.613                |
| Age and THHg                                  | 210 | 45.1 (41.8, 48.4)              | 44.2 (41.0, 47.4)    | 0.700                |
| Age and fish/meat consumption                 | 210 | 46.5 (44.1, 48.9)              | 42.9 (40.6, 45.2)    | 0.029                |
| Age, THHg and fish/meat consumption           | 210 | 46.4 (44.1, 48.8)              | 43.0 (40.7, 45.3)    | 0.036                |
| Age, THHg, fish/meat consumption, sex and SES | 186 | 46.7 (44.0, 49.5)              | 43.0 (40.4, 45.6)    | 0.035                |

Abbreviations: CI, confidence interval; SES, socio-economic status (years of parental education and family income); THHg, total hair Hg concentration; WPPSI-III, Wechsler Preschool and Primary Scale of Intelligence, 3rd edition.  
<sup>a</sup> Linear mixed effect models with a random intercept for kindergarten.  
<sup>b</sup> p-Value for the between group differences at end study.

Figs. 8.4a and 8.4b Tables including Wechsler raw scores; results from WPPSI tests, first published in Øyen et al, 2018; Kvestad et al, 2018. CC BY 4.0.

Interestingly, in the main analyses for the fish intervention trial, scientists did not find any clear-cut effects. Only when adjusting for compliance, certain subgroup analyses showed slight positive effects of fish meals on the test results. Publications have remained cautious about their findings, given the small size of the effects the obtained results showed. Caution applies as these stable test systems—such as the WPPSI test—may not be able to measure short-term changes, and these changes are difficult to interpret or are weak predictors of future cognitive capacities and functioning. Researchers are aware of these limitations and address such questions in discussion sections of their articles, yet they need to work with what is available, making the best choices for the collaborations and projects they are invited into. Comparing the intervention arms, the study found no increase in the main analyses, but small effects for single sub-analyses. Despite the work of stabilizing the validity of the study, experimental subjects also bring to the table their own preferences and tactics.

Marketing often emphasizes the general health benefit of omega-3 fatty acids in fatty fish. These considerations also play a role in a broader molecular biology of early brain development. Here, the functional role of food components and their metabolites have been studied as part of 'molecular neurocognition'. This includes the role of the nutrients eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) in early brain development and neurocognition, for which fish is an important source in the Norwegian diet, along with LC n-3 FA, vitamin D and iodine.<sup>26</sup> In risk assessments, including the toxicology of neurotoxic pollutants, researchers often face a low effect size and many other factors that also influence cognitive development, making the effects of environmental exposures difficult to prove with the epidemiological methods and study designs at hand. Here, the validity studies and the stability of the Wechsler tests across different contexts, together with a rigorous randomized trial design, stabilize and increase the credibility of the results.

Yet, fatty fish not only contains beneficial omega-3 fatty acids, but also accumulates environmental pollutants, such as persistent organic pollutants and mercury. Hence cognition enhancement through more fish rich in omega-3 may come with increased intake of potentially

---

26 See VKM Vitenskapskomiteen for Mat og Miljø 2022, p. 5.



neurotoxic pollutants. For these reasons, researchers have also investigated the question of mercury concentrations in fish meals to assess and weigh the omega-3 benefits and the toxicity of methylmercury incurred when children increase fish consumption.<sup>27</sup> The table displayed in Figure 8.4b includes values of the increased exposure to mercury after fish meals, based on analyzing mercury content in hair samples collected from participating children. This form of comparing, balancing out, and weighing different concerns is a way of figuring out health benefits and health risks. Researchers use these trials as risk-benefit analyses. In this logic, they balance gains and harm into an optimum—as much fish as possible for increasing omega-3, until the harmful effect of increased mercury content exceeds the benefits. Given other factors of influence, quantitative assessments of these concerns remain uncertain, yet the Wechsler test systems provide a validated measure without which such assessments would not be possible.

Organized around the main research question, result tables aim for a ‘one view’ at a glance—‘construction and display’, as Lorraine Daston described this epistemic device for the tables in early modern weather science.<sup>28</sup> Enacting natural histories of the air, of a disease, or of development, among other topics, such devices enable the view from a distance and the rendering of population as a statistical object that characterizes a population body that can be measured, tracked, and monitored over time.<sup>29</sup> Tables elevate the research subject to a privileged position.<sup>30</sup> Comparing risk estimates, researchers develop efforts to assess and optimize different beneficial and harmful factors into a risk-benefit analysis. In the case of statistical hypothesis tests for the nutrition trial, tables present the numerical justifications that display the modeling and weighing for managing different factors of risk and benefit. Donna Haraway called this privileged position the sciences claim among other

---

27 Ingrid Kvestad et al, ‘Fatty Fish, Hair Mercury and Cognitive Function in Norwegian Preschool Children: Results from the Randomized Controlled Trial FINS-KIDS’, *Environment International*, 121 (2018), 1098–1105, <https://doi.org/10.1016/j.envint.2018.10.022>

28 Lorraine Daston, ‘Super-Vision: Weather Watching and Table Reading in the Early Modern Royal Society and Académie Royale Des Sciences’, *Huntington Library Quarterly*, 78.2 (2015), 187–215.

29 Foucault.

30 Lorraine Daston and Peter Galison, *Objectivity* (Zone Books, 2007).

fields a 'god trick'.<sup>31</sup> Like maps in geography, tables, statistics, and their public display have played a key role in Euro-American aspirations toward science-based governance and decision-taking.

## How WPPSI Tests Reassemble Knowledge and Education

Historicizing Wechsler tests deals with moving targets, as the tests take on different roles and functions depending on whether they are used as part of healthcare provision, clinical diagnostics or studies on social work, or research on eating well in kindergartens. Once used and plugged into specific study designs, these tests have consequences for all the actors meeting in this constellation. They transform, modify, and shift things around. Like all Wechsler tests, the WPPSI tests have widely traveled, and, in this case, well beyond clinical diagnostics into other sectors, for instance nutrition science and toxicology. Within these research fields, they are pulled in as merely one element of a much larger set of established techniques for quantification of a range of variables studied alongside each other. Population studies, integrating a complexity of factors by design, also assemble experts of different kinds, who find themselves responsible for one minor part of the study. To come up with a stabilized study design, researchers from different disciplines need to negotiate the study such that it becomes and remains scientifically sound when it comes to their own field of expertise.

In this setting of a nutrition intervention trial, the Wechsler tests have moved from being a science-based diagnostic in psychology and education to a measurement device that produces new scientific data for specific research questions—such as the effects of fatty fish meals on cognitive performance. For the scientists who joined the study because they were invited as experts in psychometrics and child neuropsychology, Wechsler tests do particular and different work. Since the Wechsler tests have been validated and used over decades internationally, researchers contributing to multidisciplinary trials perceive them as something they know in terms of how the test system performs and delivers

---

31 Donna Haraway, 'Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective', *Feminist Studies*, 14.3 (1988), 575–99, <https://doi.org/10.2307/3178066>

results that can count as scientifically sound. This is characteristic of paradoxes that scientists work with, using established methods for their measurement module while being in charge of only a certain module within the larger multidisciplinary study. This condition bears its own effects and can stabilize research designs despite the paradoxes that the methods and tests contain. With scientists insisting on adherence to rigorous science, the results did not always fulfill the fish industry's expectations that science would simply prove that fish meals would enhance human performance. Especially when part of the funding comes from pharmaceutical or fish industries, following the Wechsler test protocols rigorously contributes to making the study trustworthy within science. The trials achieve this through inviting and collaborating with independent experts for all the dimensions and factors relevant to the study. As most of these studies have shown little or no association between fish lunches and cognitive performance, researchers had to manage industry expectations by being sure about their methods, even if they may not prove any health benefits, both during and after the study.

Moreover, using psychometric tests in nutritional epidemiology contributes to foregrounding biopsychosocial dimensions, which has gained momentum in public health research with children internationally. To go beyond classic epidemiological analyses of mortality and morbidity, public health researchers strived to add and target new meaningful outcomes, describing well-being, quality of life, and mental health. There is a major discussion in nutritional epidemiology about the shift to include new outcomes in addition to mortality and biomedical classifications.<sup>32</sup> Here, Wechsler tests join as inscription devices within the common conceptual framework of the multifactorial 'web of causation'.<sup>33</sup> The web of causation is the key epistemic framework of effect studies and has been at the core of statistical rationality and evidence-based reasoning. Techniques of hypothesis testing build on centuries of statistical thinking, especially on early twentieth-century biometrics, as well as on the research direction

---

32 Personal communication.

33 Nancy Krieger, 'Epidemiology and the Web of Causation: Has Anyone Seen the Spider?', *Social Science & Medicine*, 39.7 (1994), 887–903, [https://doi.org/10.1016/0277-9536\(94\)90202-X](https://doi.org/10.1016/0277-9536(94)90202-X)

of epidemiology as the study of risk factors that emerged in the 1960s and 1970s, including its study designs.<sup>34</sup> Wechsler tests are stabilized here precisely because, as a measurement and testing infrastructure, these test systems feature as scientifically sound. This is because they are already available, translated, studied, evaluated, validated, and tested numerous times, and thus have become, if not something robust, 'at least we know what it is'.<sup>35</sup> In this sense, Wechsler tests are readily available to be plugged into an epidemiological study design to address multifactorial causation, where health scientists seek tools and data to feed this web of causation and are eager to innovate toward new outcomes.

In the case of nutrition trials, the Wechsler test systems persist because they align with powerful epistemic infrastructures, which are not only those of clinical testing and neuropsychology, but also more broadly the knowledge practices within epidemiological studies of mental well-being, as well as health policymaking and health promotion in education. Starting in the early 2000s, taking their cue from the rise of evidence-based medicine, evidence-based policymaking has expanded across sectors, beyond public health, nutrition studies, and mental health research more broadly. In pedagogy, interestingly, the randomized intervention trial has a longer history compared to other fields of evidence-based policymaking. Research psychology has its own traditions of laboratories and experimental work, which have been part of the randomized trial's complex genealogy and long-standing vision of 'mental engineering'.<sup>36</sup> While evidence-based reasoning has been challenged in clinical medicine and public health, it continues to grow in importance in the policy sector, fueled by digitalization policies and a persistent push toward cost-efficiency and (rational choice) decision-taking.

For studies in public health and digital modeling, data generated in the intervention trials are forms of data extraction. Not only are they a source of data, they also feed back into the digital repositories of health

34 Alfredo Morabia, *A History of Epidemiologic Methods and Concepts* (Birkhauser, 2004).

35 Personal communication.

36 See Yerkes 1942, quoted according to James H. Capshew, J. H. 'Psychologists on site: A reconnaissance of the historiography of the laboratory', *American Psychologist*, 47(2) (1992), 132–142.

policy-making. In this vein, educational institutions become quasi-experimental environments and a real-time data source for optimizing and evaluating the effects of fish meals on cognitive performance metrics. Interestingly, salmon science, nutrition studies, and health interventions on fish and on humans are simultaneously co-optimizing both aquaculture and human performance for the ‘public good’.<sup>37</sup> ‘Seeing like an intervention trial’<sup>38</sup> for human neuro-enhancement has implications beyond knowledge production.

For educational institutions in early childhood pedagogy, conducting such studies stabilizes evidence-based health promotion as a goal. The fact that such studies are possible and deemed necessary for scientific evidence and decision-making about kindergarten diets habituates pedagogues, parents, and children to the kindergarten as a place under evidence-based optimization and as a site of knowledge production to this end. Hence, the fish trials contribute to enrolling institutions into this mode of public health governance. Tests respond to and participate in the quest for more data and use of evidence-based reasoning in pedagogy and beyond. Their calculative devices carry on their epistemic infrastructures, enacting a biopolitics of ‘populationalized’ care and control.<sup>39</sup> Effects include a focus on compliance for the sake of the higher good of the trial—at a place and among age groups who are not commonly autonomous, docile research subjects, but a vulnerable population, with consent given by parents on their behalf. Early on, being subjected to measurement and optimizing—at population levels—are routinized and learned as neutral techniques. Kindergartens as educational institutions become enrolled in optimizing toward higher performance, which presents a particular form of ‘social engineering’ prevalent in Scandinavian post-welfare states and their aspirations of industry–public-sector collaboration.

---

37 Ana Delgado and Susanne Bauer, ‘Extractions: Data Infrastructures and the Public Good’, *Science, Technology, & Human Values*, 49.3 (2024), 435–42, <https://doi.org/10.1177/01622439241242872>

38 See Crystal Biruk, ‘Seeing like a research project: producing “high-quality data” in AIDS research in Malawi’, *Medical Anthropology*, 31.4 (2012), 347–66.

39 Christine Holmberg, Christine Bischof, and Susanne Bauer, ‘Making Predictions: Computing Populations’, *Science, Technology, & Human Values*, 38.3 (2013), 398–420, <https://doi.org/10.1177/0162243912439610>

## Coda: Wechsler Tests as Plug-in Modules

This chapter has been an attempt to render visible the ways in which WPPSI tests work to reassemble actions, tests, participants, institutions, and society. By connecting public health messages on eating fish with human brain development, institutionalized research on nutrition and cognitive abilities routinize practices of optimization from an early age, including school and kindergarten settings. These knowledge practices—intervention studies in educational settings—habituate public institutions to specific modes of evaluating and optimizing. Together with the research on metabolism, nutrition, and fish diets, these nutrition intervention studies “co-experimentalize” kindergarten lunches, educational institutions, and the ability to perform. This actively shapes how the relations between fish meals rich in omega-3 and the goal of speeding up cognitive performance and enhancing abilities are conceived and achieved.

At first glance surprisingly, Wechsler tests also feature, in addition to diagnostic testing, in areas that are more distant from clinical applications—in research settings such as nutrition studies or environmental health. One main goal of this chapter has been to better understand how and why kindergarten kids in Bergen were made to complete intelligence tests while subjected to different diets. Following the work of Wechsler tests in such a setting has shown how and why these intelligence tests have been so persistent. They continue to be present and evolve as scientific objects and perform various social roles in today's Norway. Wechsler tests have traveled a long way and, precisely because of their trajectories, continue to perform as well-known standards trusted to do certain work. In our case of an intervention trial in nutritional epidemiology, Wechsler tests are picked up, plugged into a study design as a module, and combined with other tests and metrics while adhering to scientific methodologies of the RCT, tailored to specific research questions. Producing knowledge becomes about putting together the right boxes for the job, the WPPSI being one of them. While part of a broader agenda of studies about well-being and mental health in relation to diet, these practices feed into normative ableist projects around “being smart”. This contributes to further stabilizing and institutionalizing Wechsler IQ tests as routinized tools of

expert practitioners to be used for collective optimization projects.

Norway's science policy, envisioning scientific research to solve what policy has formulated as society's "grand challenges", has prompted its own modes of optimization in an era of datafication. The Wechsler tests, as one of the very few standardized, comprehensively validated test systems normed for the Norwegian context, contributes to assembling scientists, the state, and the aquaculture industry, feeding into and translating their respective demands. In turn, such usage of the Wechsler test systems further enrolls institutions, parents, participating children, and media. Despite some friction, they all aligned in temporarily transforming some kindergartens into controlled environments of knowledge-making for the Norwegian (post-)welfare state—and the industries it lives by. In that sense, it should not surprise us that the test continues to persist and travel, especially given the ongoing calls for datafication and digital extraction.<sup>40</sup>

The broad usage of Wechsler tests in Scandinavian public education systems and diagnostics is different from commercialized consulting activities around the WPPSI tests widely offered in the US through sites such as Testing.Mom.com,<sup>41</sup> which demonstrates how people work to prepare for these tests in order to secure access to good schools. In the Nordic public education systems, the employment of intelligence testing is more institutionally embedded and is part of a wider mode of evidence-based optimization. Here, intelligence tests are part of larger batteries of testing devices, which are used when public actors or policymakers call on science to solve problems, from the fish industry to public policy. The WPPSI test usages examined in this case-study function in a double sense. Firstly, using such established and validated tests assures that researchers use scientifically sound instruments; and secondly, researchers chose these tests as "best tools" at hand to deliver the modules within the commissioned study they were asked to contribute to. As scientifically sound modules, they can enroll scientists

40 Delgado and Bauer; Susanne Bauer, 'Seeing Like a Model Fish: How Digital Extractions Mediate Metabolic Relations', *Science, Technology, & Human Values*, 49.3 (2024), 524–54, <https://doi.org/10.1177/01622439241236520>.

41 These private services offer preparation workbooks for children with artwork and tasks, similar to the actual tests, advertising to parents the 'most economical options' for private school admission preparation. See for instance: <https://www.youtube.com/watch?v=Z3r-UCnM5UU>

into projects that are often not of their own choosing. By holding on to the most established and validated methods, they might advocate and use the tests within rigorous methodologies strategically, at times as protection from industry influence. They are part of scientists' response to calls from science policymakers in Norway, who ask researchers to address "grand challenges" and contribute to aquaculture-centred bioeconomy aspirations, combined with evidence-based governance of public health and education. These applications in practice, in turn, further stabilize the usage of IQ tests as a research tool.

This chapter has shown how the Wechsler tests have entered into nutrition research as validated measurement devices. Studies of cognitive performance through optimized fish diets show how public health—not only in the post-welfare state—has continued to be embedded in the fostering of pre-defined abilities. Part of health promotion, they enact an ableist project of children educated in order to meet the conditions for cognitive high performance. Providing conditions for enabling better performance has become part of pedagogy, care, and the public good. This feature has been characteristic of social policy and the welfare state from its beginning, as a state measure to maintain the workforce, rather than a project of reducing health inequalities. With health promotion as common sense in public health, several paradoxes often remain. Firstly, the epistemic infrastructures of intervention trials combine different modules that are plugged into the trial's architecture. For common clinical and diagnostic purposes, test manuals advice against detecting changes at short intervals. Yet, researchers still performed tests at short intervals to produce data for an intervention trial—without the goal of an individual diagnosis, but rather as a routine measurement device to generate data for group comparisons before and after different nutrition interventions to record potential changes. Secondly, while present in the public debate, inclusion and neurodiversity have not entered the registers and categories that inform intervention trials in public health, which often still focus on measuring and improving performance. It is here that an implicit adherence to the desirability of the dimensions measured in intelligence tests remains present, a form of in-built ableism.<sup>42</sup> Rather, their epistemic frameworks shaped in the twentieth

---

42 Sunaura Taylor, *Beasts of Burden. Animal and Disability Liberation* (New York: The New Press 2017).



century have remained unaddressed and unchanged, as they help scientists to navigate and distance themselves from industry interests. Because of its many validations, the stability of the WPPSI test has rarely been challenged, and each plugging-in of Wechsler tests to an empirical study in public health, nutrition or education contributes to the tests' mobility across fields and functions, adding to their overall persistence.

## Bibliography

- Adams, Vincanne, Michelle Murphy, and Adele E Clarke, 'Anticipation: Technoscience, Life, Affect, Temporality', *Subjectivity*, 28.1 (2009), 246–65, <https://doi.org/10.1057/sub.2009.18>
- Asdal, Kristin and Tone Huse *Nature-Made Economy: Cod, Capital and the Great Economization of the Ocean* (Cambridge MA: MIT Press, 2023).
- Bauer, Susanne, 'Seeing Like a Model Fish: How Digital Extractions Mediate Metabolic Relations', *Science, Technology, & Human Values*, 49.3 (2024), 524–54, <https://doi.org/10.1177/01622439241236520>
- Bauer, Susanne, Martina Schlünder, and Maria Rentetzi, *Boxes. A Field Guide* (Manchester: Mattering Press, 2020).
- Biruk, Crystal, 'Seeing like a research project: producing "high-quality data" in AIDS research in Malawi', *Medical Anthropology*, 31.4 (2012), 347–66, <https://doi.org/10.1080/01459740.2011.631960>
- Bowker, Geoffrey C. and Susan Leigh Star, 'Sorting Things out: Classification and Its Consequences', *Inside Technology* (Cambridge MA: MIT Press, 1999).
- Buchardt, M., P. Markkola, and H. Valtonen, 'Introduction: Education and the Making of the Nordic Welfare States', in *Education, State and Citizenship* (University of Helsinki, 2013), pp. 7–30.
- Bugge, Markus, Teis Hansen and Antje Klitkou, 'What Is the Bioeconomy?', in *From Waste to Value: Valorisation Pathways for Organic Waste Streams in Circular Bioeconomies*, ed. by Antje Klitkou, Arne Martin Fevolden, and Marco Capasso (London: Routledge, 2019), pp. 19–50.
- Campbell, Fiona Kumari, *Contours of Ableism: The Production of Disability and Ableness* (New York: Palgrave Macmillan, 2009).
- Carson, John, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750–1940* (Princeton NJ: Princeton University Press, 2006).
- Daston, Lorraine, 'Super-Vision: Weather Watching and Table Reading in the Early Modern Royal Society and Académie Royale Des Sciences', *Huntington Library Quarterly*, 78.2 (2015), 187–215.

- Daston, Lorraine, and Peter Galison, *Objectivity* (New York: Zone Books, 2007).
- Delgado, Ana and Susanne Bauer, 'Extractions: Data Infrastructures and the Public Good', *Science, Technology, & Human Values*, 49.3 (2024), 435–42, <https://doi.org/10.1177/01622439241242872>
- Desrosières, Alain, *The Politics of Large Numbers: A History of Statistical Reasoning* (Cambridge MA: Harvard University Press, 1998).
- Foucault, Michel, *The Birth of Biopolitics: Lectures at the Collège de France, 1978–1979*, trans. by G. Burchell (New York: Palgrave Macmillan, 2008).
- Hamre, Bjørn, Thom Axelsson, and Kari Ludvigsen, 'Psychiatry in the Sorting of Schoolchildren in Scandinavia 1920–1950: IQ Testing, Child Guidance Clinics, and Hospitalisation', *Paedagogica Historica*, 55.3 (2019), 391–415, <https://doi.org/10.1080/00309230.2019.1597137>
- Haraway, Donna, 'Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective', *Feminist Studies*, 14.3 (1988), 575–99, <https://doi.org/10.2307/3178066>
- Holmberg, Christine, Christine Bischof, and Susanne Bauer, 'Making Predictions: Computing Populations', *Science, Technology, & Human Values*, 38.3 (2013), 398–420, <https://doi.org/10.1177/0162243912439610>
- Krieger, Nancy, 'Epidemiology and the Web of Causation: Has Anyone Seen the Spider?', *Social Science & Medicine*, 39.7 (1994), 887–903, [https://doi.org/10.1016/0277-9536\(94\)90202-X](https://doi.org/10.1016/0277-9536(94)90202-X)
- Kvestad, Ingrid, Silje Vabø, Marian Kjellekvold, et al, 'Fatty Fish, Hair Mercury and Cognitive Function in Norwegian Preschool Children: Results from the Randomized Controlled Trial FINS-KIDS', *Environment International*, 121 (2018), 1098–1105, <https://doi.org/10.1016/j.envint.2018.10.022>
- Margetts, Barrie, Daniel Warm, Agneta Yngve, and Michael Sjöström, 'Developing an Evidence-Based Approach to Public Health Nutrition: Translating Evidence into Policy', *Public Health Nutrition*, 4.6a (2001), 1393–97, <https://doi.org/10.1079/PHN2001226>
- Mol, Annemarie, Ingunn Moser, and Jeannette Pols, eds., *Care in Practice, On Tinkering in Clinics, Homes and Farms* (Bielefeld: transcript Verlag, 2010), <https://doi.org/10.1515/transcript.9783839414477>
- Morabia, Alfredo, *A History of Epidemiologic Methods and Concepts* (Basel: Birkhauser, 2004).
- Øyen, Jannike, Ingrid Kvestad, Lisa Kolden Midtbø, et al, 'Fatty Fish Intake and Cognitive Function: FINS-KIDS, a Randomized Controlled Trial in Preschool Children', *BMC Medicine*, 16.1 (2018), p. 41, <https://doi.org/10.1186/s12916-018-1020-z>
- Porter, Theodore M., *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton NJ: Princeton University Press 1995).

- Sackett, David L., W. Scott Richardson, William Rosenberg, and R. Brian Haynes, *Evidence Based Medicine* (Churchill Livingston, 1995).
- Stigler, Steven M., *The History of Statistics: The Measurement of Uncertainty before 1900* (Cambridge MA: Belknap Press of Harvard University Press, 1986).
- Taylor, Sunaura, *Beasts of Burden. Animal and Disability Liberation* (New York: The New Press, 2017).
- Tröhler, Daniel, 'The Medicalization of Current Educational Research and Its Effects on Education Policy and School Reforms', *Discourse: Studies in the Cultural Politics of Education*, 36.5 (2015), 749–64, <https://doi.org/10.1080/01596306.2014.942957>
- Vallgård, Signild, 'Governing People's Lives: Strategies for Improving the Health of the Nations in England, Denmark, Norway and Sweden', *European Journal of Public Health*, 11.4 (2001), pp. 386–92, <https://doi.org/10.1093/eurpub/11.4.386>
- Verran, Helen, 'Decomposing Numbers', *HAU Journal of Ethnographic Theory*, 8.1–2 (2018), 23–26.
- Verran, Helen, 'Re-Imagining Land Ownership in Australia', *Postcolonial Studies*, 1.2 (1998), 237–54, <https://doi.org/10.1080/13688799890165>
- Vrangbæk, Karsten, 'The Regulation of Health Care in Scandinavia', in *Professional Health Regulation in the Public Interest*, ed. by John Martyn Chamberlain, Mike Dent, and Mike Saks, International Perspectives, 1st edn. (Bristol: Bristol University Press, 2018), pp. 61–76, <https://doi.org/10.2307/j.ctv1xxs5q.10>
- Ydesen, Christian, 'Crafting the English Welfare State—Interventions by Birmingham Local Education Authorities, 1948–1963', *British Educational Research Journal*, 42.4 (2016), 614–30, <https://doi.org/10.1002/berj.3223>

## 9. A Norwegian Terman–Merrill and the Shelf Life of Intelligence Tests

*Gard Paulsen*

---

In the late autumn of 1954, the publisher J. W. Cappelen issued the Norwegian translation and adaptation of the Stanford–Binet intelligence test, *The Norwegian Standard of Terman–Merrill’s Revision of the Binet Tests*.<sup>1</sup> The title, almost maze-like in its intricacies, revealed nodes in the circulation of intelligence tests—the American original of the Norwegian adaptation was first published in 1937.<sup>2</sup> It was made by Lewis Terman and Maud Merrill, both of Stanford University, and published by the American publishing company Houghton Mifflin. It was a revision of an earlier translation and adaptation of the original French Binet test, first made by Terman in 1917. The Norwegian translation was long in the making, but when it was finally published and made available in a slim volume, the test stuck and became widely used. A second printing was made as late as 1970. The publisher also printed the record booklets repeatedly. The test’s shelf life proved far beyond what was expected upon its introduction. A similar longevity appears to be a distinctive feature of many psychological tests.<sup>3</sup>

This chapter approaches the shelf life of intelligence tests by analyzing how the Terman–Merrill revision of the Stanford–Binet was adapted and utilized in Norway during the postwar period. Considering the

---

1 Johs. Sandven, *Norsk standard av Terman–Merrills Stanford-revisjon av Binet-prøvene* (Cappelen, 1954).

2 Lewis M. Terman and Maud A. Merrill, *Measuring Intelligence: A Guide to the Administration of the New Revised Stanford-Binet Tests of Intelligence* (Houghton Mifflin Company, 1937).

3 Ken Richardson, *Understanding Intelligence* (Cambridge University Press, 2022).

shelf life of intelligence tests and their formation as a product traded and sold, if not entirely “off the shelf”, but through a specialized trade, this chapter reveals a much-overlooked aspect in the history of psychology by attending to the commercial concerns of psychological tests.<sup>4</sup> Furthermore, this chapter moves beyond arguments regarding professional preference or scientific sophistication when considering the longevity of specific tests, highlighting in particular how they were adapted to and shaped by the local context and market of Norway. Drawing on original archival sources from publishers, translators, authors, researchers, numerous research reports, record booklets, and tests, the analysis utilizes material that has often remained unexamined, disregarded, or overlooked in the historical literature.<sup>5</sup>

## Longevity, Materiality and Commercial Availability

“Shelf life” typically refers to the period during which a commodity can be stored without becoming unfit for use, consumption, or sale. Not a regular commodity, the expiry date for intelligence tests seems almost indefinite, and indeed, some have stayed on the shelves for a very long time. Even the Stanford–Binet is a product still sold and used today, although it differs significantly from the original versions. If one considers the history of specific test items, the longevity of what makes up an intelligence test becomes even more striking.<sup>6</sup>

The longevity of intelligence tests, however, comes with adaptations and revisions of some of the most widely used tests and test items. Numerous iterations and combinations of tests and test items mark the history of intelligence tests—for example, the aforementioned Norwegian 1954 adaptation of the Terman–Merrill 1937 test was, in turn,

---

4 On calls for a renewed interest in the history of intelligence tests, see Annette Mülberger, ‘Biographies of a Scientific Subject: The Intelligence Test’, in *Oxford Research Encyclopedia of Psychology*, by Annette Mülberger (Oxford University Press, 2020); Annette Mülberger, ‘The Need for Contextual Approaches to the History of Mental Testing’, *History of Psychology*, 17.3 (2014), 177–86.

5 See the appendix for details about the archival material used in this chapter.

6 For example, digit spans, well, span the entire history of intelligence testing. Despite numerous claims about its insignificance and unsuitability, there is no getting rid of it. See Corwin Boake, ‘From the Binet–Simon to the Wechsler–Bellevue: Tracing the History of Intelligence Testing’, *Journal of Clinical and Experimental Neuropsychology*, 24.3 (2002), 383–405.

a revision of Terman's 1916 revision of the Binet original. It is easy to get lost among the various versions.<sup>7</sup> In the 1950s, the Norwegian publisher considered the genealogy of the Terman–Merrill intelligence test so complex that it seemed impossible to give the Norwegian adaptation a proper name. To envision how it could be cataloged and indexed by libraries and booksellers even caused anxiety among the editors. The list of the test's figurative great-grandparents, grandparents, parents, and godparents—as an editor at the Norwegian publisher formulated it—was so long that the editor doubted he could fit it all on a page. The editor was even concerned about the validity of the term finally chosen for the title, 'The Norwegian Standard', fearing protests from the Norwegian organization for industrial standardization.<sup>8</sup>

The Stanford–Binet of today is far removed from the adaptations issued by the Norwegian publishers in the early 1950s. Just as there was a need to adapt a test to a local context, language, and culture, there was a regular but slow-moving need to bring the test 'up to date', a term Maud Merrill applied in 1960 when she argued for a revision of the original 1937 US edition.<sup>9</sup> However, the need to bring tests "up to date" has implied different things at different times; the presence of obsolete materials has regularly triggered the need for changes.<sup>10</sup> New statistical techniques have also been introduced to justify and alter tests, and theoretical changes have influenced how tests are structured and interpreted.

Despite the variations, the Stanford–Binet should, like all other intelligence tests, primarily be understood as a 'material, procedural system', as historian Peter Galison interprets the history of the Rorschach test.<sup>11</sup> Undoubtedly, the test possessed a formality, which certainly mitigated against extensive changes and alterations. In the original

7 Lewis M. Terman and Maud A. Merrill, *Stanford–Binet Intelligence Scale; Manual for the Third Revision Form L–M* (Houghton Mifflin Company, 1960).

8 The organization in question was the Norsk standardiseringsforbund. See J. W. Cappelen to Johs. Sandven, 6 May 1954, correspondence in the archive after J. W. Cappelen, held by the National Library of Norway, Oslo (hereafter abbreviated to JWC NB).

9 Terman and Merrill, *Stanford–Binet Intelligence Scale; Manual for the Third Revision Form L–M*, p. 19.

10 Terman and Merrill, *Stanford–Binet Intelligence Scale; Manual for the Third Revision Form L–M*, p. 20.

11 Peter Galison, 'Image of Self', in *Things That Talk: Object Lessons from Art and Science*, ed. by Lorraine Daston (Zone Books, 2004), pp. 257–94 (p. 258).

edition of the Terman–Merrill revision, the test items and the scoring procedure were detailed in a published volume. The test items were also spelled out in the published volume of the Norwegian adaptation, including instructions regarding test procedures and scoring. Here, the regimented testing phases were outlined and how to consider the responses was specified, as the test came with a list of correct answers and examples of what should be regarded as incorrect. It dictated how to proceed with the scoring. The instructions in the manual, the phrasing of the different questions and tests, and the record booklet all contributed to what was considered the proper way of testing. These testing regiments could be understood as being imbued with a particular materiality. In Cornelia Vismann’s now classic *Akten: Medientechnik und Recht*, she studied the concrete techniques by which law was processed, in particular using files. She argued that law and authority ‘assume concrete shape in files’.<sup>12</sup> Intelligence, I suggest, is of a similar nature: it lies in the processing of tests more than in psychological theory. Furthermore, intelligence tests also embedded procedures and elements that afforded their inclusion in formal filing systems similar to those of Vismann’s study. For example, the American original came with a pre-printed abbreviated card file, and the record booklet accompanying the Norwegian edition was prepared with the first page already set for the inclusion of aggregated results and personal information. As such, intelligence was in the processing of files, just as much as in the processing of tests.

However, the procedural system of the test also hinged on a different kind of materiality: the many physical objects needed for the 1937 Stanford–Binet intelligence test were housed in a wooden box with a sliding lid. These were not mere packaging. Instead, tests often relied on packaging to lend credibility and facilitate portability, both practically and institutionally. According to the historian John Carson, the commodification and packaging of tests in the interwar period were at the heart of the rapid dissemination of intelligence testing in American society.<sup>13</sup>

12 From the English abridged translation of the original is Cornelia Vismann, *Files: Law and Media Technology*, Meridian (Stanford University Press, 2008), p. xiii.

13 John Carson, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750–1940* (Princeton University Press, 2007), pp. 253–54. Recently, attention to the packaging of the tests has been reiterated in a novel

Nevertheless, the Norwegian version was officially only the bound volume. However, a leather suitcase was sometimes used to store all the various test items needed to hold a complete set.<sup>14</sup>

The statistical policies used in the original test also proved to be very malleable. The term ‘standardized’, or as in the Norwegian adaptation, ‘the Norwegian standard’, has typically been used to indicate various formal adaptations, in particular, statistical adjustments. However, the different statistical procedures used in the different tests varied to the extent that a standardized test did not necessarily imply the use of a normalized score with standard deviations. For example, when revising the Stanford–Binet in 1937, ‘no attempt was made [...] to secure a normal distribution of the resulting IQs. Items were chosen so that the average mental age for an age group coincided with their chronological age’.<sup>15</sup> A similar policy was adopted when adjusting for the Norwegian context. In the following, I use the term ‘statistical adjustments’ for work concerned with the statistical properties and design of the tests, even though the term ‘standardization’ is generally used to imply any kind of statistical adaptation of tests. However, this did not mean the use of normalized scores, standard deviations, and so on, neither in the original design of the test nor in the adaptation.

Over time, the part of a test that has been deemed immutable or understood as changeable has often been shaped by the concerns and interests of test publishers and funding agencies. Concerns and considerations regarding distribution logistics have also regularly influenced how intelligence tests are administered. Understanding the interest of publishers and distributors in shaping the final make-up of tests is a concern this chapter shares with book history.<sup>16</sup>

Despite the rigidity of the procedural system that is the intelligence

---

approach to the history of intelligence testing. For this, see Alexandra Bergstrom-Katz, ‘On Intelligence Tests: Psychological Objects and Their Subjects’ (Unpublished PhD, Department of Psychosocial Studies, Birkbeck, University of London, 2023); Jacy L. Young, ‘Test or Toy? Materiality and the Measurement of Infant Intelligence’, *History of Psychology*, 18.2 (2015), p. 103.

14 See Fig. 9.3 in this chapter.

15 Quinn McNemar, *The Revision of the Stanford–Binet Scale: An Analysis of the Standardization Data* (Houghton Mifflin, 1942), p. 17.

16 On the importance of publishers in recent book history, see for example Robert Darnton, ‘“What Is the History of Books?” Revisited’, *Modern Intellectual History*, 4.3 (2007), 495–508; James Raven, *Publishing Business in Eighteenth-Century England* (The Boydell Press, 2014).



test—understood both formally, statistically, and materially—changes and alterations were being made, particularly as tests were adapted to new contexts and locations. In the following, I will survey the longevity of the Stanford–Binet by exploring the degrees of immutability and flexibility at play when the 1937 Terman–Merrill was adapted for use in Norway. Firstly, a brief overview of the availability and distribution of intelligence tests more generally is provided.

## Tests that Travel

In 1908, the American psychologist Henry Herbert Goddard toured Europe and learned about the intelligence tests constructed by Alfred Binet and Théodore Simon.<sup>17</sup> He soon began to use his translation of the test at the Vineland Training School of New Jersey, where he headed a psychological laboratory.<sup>18</sup> Eventually, Goddard obtained Simon's permission to translate and distribute a version of the test. From 1910 to 1916, Goddard's operations at Vineland distributed a print run of 22,000 pamphlets and 88,000 record booklets of a very condensed version of the Binet test, all 'without effort or advertisement'.<sup>19</sup> In 1916, Goddard also published the test as a literal translation of the French original as part of a series of publications by the training school and the small medical publisher Williams and Wilkins Company.<sup>20</sup>

Many new intelligence tests were soon published in the United States after Goddard's efforts. Lewis Terman's Stanford–Binet revision, published in 1916, received great prominence, and was at the time considered the most thorough of all the tests based on the French original. It followed Goddard's lead by issuing one main text and a separate record booklet that would facilitate keeping the scores and recording individual answers. The leading publication was a combination of instructions, details of test items, and an in-depth discussion of statistical

17 James Trent, *Inventing the Feeble Mind: A History of Intellectual Disability in the United States* (Oxford University Press, 2016); Leila Zenderland, *Measuring Minds: Henry Herbert Goddard and the Origins of American Intelligence Testing* (Cambridge University Press, 1998); Carson.

18 See Goddard's own account of this in the introduction of *Alfred Binet and Théodore Simon, The Development of Intelligence in Children*, trans. by Elizabeth S. Kite (Williams & Wilkins company, 1916).

19 Binet and Simon, p. 6.

20 Binet and Simon.

evidence in the format of a textbook put out by The Riverside Press, an imprint of the Houghton Mifflin Company. Its companion volume dealt with the statistical attributes of the test.<sup>21</sup> A more informal user guide for schoolteachers was published in 1919, again authored by Terman and published by Houghton Mifflin.<sup>22</sup> With all this, intelligence tests had reached the mainstream of America and of American publishing, as they moved beyond outlets like Goddard's self-distribution or professional journals. With Houghton Mifflin, the intelligence tests also reached the limitations set by American copyright. By 1922, the record booklet of the Stanford–Binet, for example, warned that any form of reproduction of its contents was considered a violation of copyright.<sup>23</sup>

Whereas Goddard's efforts were networked and personal, including training teachers at a specialized summer school at Vineland, Terman scaled up and branched out, to such an extent that his biographer argued that Terman advanced test publishing in general.<sup>24</sup> Besides publishing the Stanford–Binet, this included influence on the publishing of group tests intended for large-scale testing in public education. Several American publishers were involved in various tests, often targeting slightly different markets. While the first commercial group test, the Otis Group Intelligence Scale, was published in 1919 by the World Book Company, by 1929, several million tests were being administered annually in public schools by many publishing houses.<sup>25</sup>

Eventually, Terman's tests were exported—some also officially approved by himself, just as his revision carried an approval seal by the remaining author of the French original, Théodore Simon. The first exports of the Terman test were exact copies, such as the one published by the British publisher George G. Harrap in 1919. Another iteration of this was published in the 1930s in another Harrap-published version: 'The alterations made in this edition are chiefly verbal, the more important

21 Lewis M. Terman, *The Stanford Revision and Extension of the Binet–Simon Scale for Measuring Intelligence* (Warwick & York, Inc., 1917).

22 Lewis M. Terman, *The Intelligence of Schoolchildren: How Children Differ in Ability. The Use of Mental Tests in School Grading and the Proper Education of Exceptional Children* (Houghton Mifflin, 1919).

23 See both editions held, for example, by the National Museum of American History.

24 Henry L. Minton, *Lewis M. Terman: Pioneer in Psychological Testing* (New York University Press, 1988), pp. 107–9.

25 JoAnne Brown, *The Definition of a Profession: The Authority of Metaphor in the History of Intelligence Testing, 1890–1930* (Princeton University Press, 1992), p. 110.

of these having been submitted to and approved by Professor Terman'.<sup>26</sup>

By the early 1930s, the American test markets seemed so sizable that the young psychologist Oscar Krisen Buros intended to set up what he described as a 'test consumers' research organization', inspired by the establishment of the general testing and rating agency of Consumers' Research a year before. While failing to secure financial support for such an operation, Buros was able to set up a publishing operation alongside his career as an academic psychologist at Rutgers University. By 1933, a bibliography comprising 257 new tests was published, revised or supplemented in 1933 and 1934. By 1938, the first of many *Mental Measurement Yearbooks* was published, containing information about published tests and test reviews written by specialists.<sup>27</sup>

In the first edition of Buros' yearbook, one of the tests reviewed was the revised edition of Terman's Stanford-Binet, published in 1937. Ten years into the project, the revision resulted from extensive trials and data gathering, to the extent that a Hollerith tabulation machine—a precursor to modern computers—and punch cards were used to sort the scores. The *Yearbook* remarked that the publication was considered 'a genuine cultural event'.<sup>28</sup> Quite a lot had changed in form and conceptualization. No longer was the test accompanied solely by books and the record booklet, but two wooden cases full of props, toy-like miniatures, and printed cards were issued to accompany the greatly expanded test. Indeed, with the collaboration of psychologist Maud Merrill, the test now included two sets of test items and scoring scales.

## Norwegian Adaptations

In 1938, the Norwegian psychologist Helga Eng obtained one of the two elaborate boxes of the recently published revision of the Stanford-Binet tests, alongside the published instructions in Terman and Merrill's *Measuring Intelligence* textbook. For Eng, the box was hardly the most significant event of the year, as she became the first chair of education

26 See the George G. Harrap editions of the two Terman books.

27 The operation has continued up until today, run from The Buros Centre, an independent organization within the University of Nebraska-Lincoln. Oscar K Buros, 'Fifty Years in Testing: Some Reminiscences, Criticisms, and Suggestions', *Educational Researcher*, 6.7 (1977), 9–15.

28 Oscar K. Buros, ed., *The 1938 Mental Measurements Yearbook*.



The Terman–Merrill test she had obtained was purchased alongside other tests and material at Eng’s request to set up the Department of Education at the University of Oslo. Eng was no stranger to psychological tests; her interests had first been directed towards German experimental psychology. Within this tradition, she published her most influential research, which was also translated and published in German, drawing interest from prominent experimental psychologists.<sup>32</sup> However, her methodological and theoretical interests quickly broadened, encompassing the interpretation of children’s drawings and the development and construction of various assessment tools. Early on, Eng also translated tests into Norwegian. She even attempted to translate the original Stanford–Binet in 1923 and tried out her tests on a sample of eight-, nine- and ten-year-olds at a local school in Oslo.<sup>33</sup> This test ended up being partly translated by Eng’s brother in the late 1920s. In 1925, Eng also designed a group test for assessing general mental abilities, but she carefully avoided calling it an intelligence test. The test was used as a part of the various entrance tests for the vocational training school.<sup>34</sup> Over the 1920s and 30s, she also built various specific psychological tests, primarily for vocational purposes and as part of her position at the Psychotechnical Institute in Oslo.<sup>35</sup> Even though both her group test and the specialized vocational tests embedded test items that could already be found in similar and already distributed tests, she would argue that the tests were all her own creation and considered them original contributions.<sup>36</sup>

Eng’s mailbox was also regularly filled with what was typically termed a ‘specimen set’, or samples of commercially available tests. Over the 1920s and 1930s, in particular, many tests published by the

---

32 Lønnå.

33 Lønnå, p. 146.

34 Eng would later refer to the group test as ‘Gruppetest 1925’. It was detailed and criticized in Johs. Clausen, ‘Intelligensmåling i grupper’, *Norsk Pedagogisk Tidsskrift*, 25 (1941), 65–82. On Eng’s insistence on the test not being an intelligence test, see her notes to H. Witzøe, titled ‘ad H. Eng til H. Witzø’, undated, in Helga Eng papers, National library of Norway (Hereafter shortened as HE, NB). Eng has also detailed the use of this group test in Eng, *Norsk psykoteknikk i ti år 1925–1935*.

35 An overview is given by Emil Østlyngen, *Aktuelle tester i Norge* (Norsk Psykologforening, 1957); Emil Østlyngen, *Bibliografi over psykologisk faglitteratur av norske forfattere* (Norsk psykologforening, 1942).

36 Eng accounted for most of her test design, in retrospect, in a summary letter to Emil Østlyngen, Oslo, 6 January 1957, HE, NB.

American World Book Company as well as German publishers were obtained by Eng.<sup>37</sup> None of the tests Eng developed herself were made available commercially, although they were used extensively at the Psychotechnical Institute and the vocational training school.<sup>38</sup> To the extent they were distributed beyond her use, it was at her own expense—in the 1930s, she even paid most of the costs for distributing 1,000 copies of her group test to a local schoolteacher.<sup>39</sup> Indeed, Eng preferred to keep the tests a secret, as she believed they would lose their value as entrance tests at the vocational training school if they were published in any sense.<sup>40</sup>

Eng was no exception in this. There were few Norwegian tests available for general distribution or sale. The psychologist Martin L. Reymert had made a rudimentary translation and adaptation of the original Stanford–Binet available for purchase in 1924. The test, however, was privately printed and distributed, rather than being carried by a publisher or bookstore. Reymert followed the original made by Lewis Terman closely, even producing a limited print run of record booklets and making the illustrations available for sale. Reymert, however, emigrated to the US himself soon after.<sup>41</sup> As such, the test was only available in a brief fire sale before he moved abroad.<sup>42</sup> Another Norwegian adaptation of the original Stanford–Binet was created by Johan Lofthus in 1931. The work was his doctoral thesis and was published “in commission” by the publisher Aschehoug.<sup>43</sup> This was, however, a consequence of the regular publication of all doctoral theses approved by the University of Oslo. No new editions were released when the original and limited print run had sold out.

Over the 1930s, two translated and standardized group tests were

37 The collection of “specimen kits” and sample tests collected by Eng is now held by the Department of Education, University of Oslo. See catalog on [www.alvin-portal.org](http://www.alvin-portal.org)

38 Her initiative of distribution, carrying a substantial amount of the cost herself, is mentioned in letters to H. Witzoe, Kristiansand, March and April 1937, in HE, NB.

39 See above. See also the receipt for the printing in 1937, from the O. Fredr. Arnesen printer, in HE, NB.

40 On her reluctance to share the group test, see letter from Eng to K. Mønnesland, Oslo, 18 September 1938, HE, NB.

41 On Reymert and his adaptation of the Stanford–Binet, see chapter 1, this volume.

42 *Tidsskrift for Den norske legeforening*, vol. 45, nr. 11, 598.

43 On Lofthus and his adaptation of the Stanford–Binet, see chapter 1, this volume.

available in Norway.<sup>44</sup> The American Dearborn test, first published in the US by J. B. Lippincott of New York, was translated and statistically adjusted on a small sample of children in Trondheim by Marie Pedersen. In 1933, the Norwegian publisher Gyldendal published the result. By 1937, Gyldendal republished the Dearborn test after further statistical adjustments based on trials in Oslo. Both Dearborn editions were limited to pupils at the entry level of Norwegian schools.

Gyldendal was also responsible for the publication of what was known as the Kuhlman and Anderson's group test, also of American origin, in 1941. This publication was notable because the American copyright holder, the Educational Test Bureau, explicitly approved the Norwegian version. We can also assume that this was the first time in Norway that such an agreement was made between two professional organizational units rather than individuals. No similar approval was printed for any of the Dearborn publications.<sup>45</sup>

While the Dearborn test was sold "above the counter" at regular Norwegian bookshops, the distribution of the Kuhlman-Anderson edition was more restrictive, limited to those who applied for a copy to the committee responsible for financing the adaptation.<sup>46</sup>

Upon assuming the chair of education in 1938, Eng began translating the instructions of the Terman-Merrill revision of the Stanford-Binet. She included the test in an official course that was eventually established at the university. She sought out other participants and allies among her students and former colleagues, including preschool teachers, schoolteachers, and principals, who collaborated with her on testing the translated version. However, the scale of the trials was minimal.<sup>47</sup>

44 Lina Bryn Larsen, 'med maaling følger viden': *Intelligensmåling i den norske folkeskolen, 1900–1945, Hovedfagsoppgave i Historie* (Universitetet i Bergen, 2006).

45 Marie Pedersen, *Intelligensprøving av barn: Dearborns gruppetest* (Gyldendal, 1933); Johan Lofthus, Augusta Rasmussen, and Bernhard Ribsskog, *Evnemålinger i folkeskolens 1. klasse: Dearborns gruppeprøve* (Gyldendal, 1937); Bernhard Ribsskog, *Evnepørve for 1. klasse i folkeskolen: standardisert på grunnlag av Kuhlmann-Andersons gruppeprøve* (Gyldendal, 1941).

46 Larsen.

47 See for example Helga Eng to Kathrine Simonsen, 27 March 1941, in folder marked 'Svar til studenter, 1938–1945', PFI UiO. Ella Esp, a nursery school teacher and child welfare inspector, was also recruited to perform some tests. See Helga Eng to Elfrida Høigård, Oslo, 6 February 1945, PFI UiO. See also Helga Eng's summary of the educational studies at the University of Oslo in *Norsk pedagogisk årbok 1942/43*, vol. 11, pp. 37–38.

Eng also initiated more extensive studies of intelligence tests in general during the very beginning of organized educational research at the University of Oslo. This programme was extended and continued after her retirement, particularly by her former student and successor as professor, Johannes Sandven.<sup>48</sup> While Eng's translation and adaptation initiative began before World War Two and continued even as the war progressed, it was only taken on in full after 1945.<sup>49</sup> There is some evidence that Eng herself tested pupils in Oslo during the war.<sup>50</sup> When the war ended, the trials and test adaptation involved several of Eng's graduate students. The work was organized so that one student was tasked with getting samples from a given age group and conducting trials of Eng's translation.<sup>51</sup> Eng also circulated her preliminary work and the results of the individual student trials to former students responsible for testing at the local school level. While the gathering of statistics by students was confined to children in Oslo, the distribution of the test material to former students across Norway even led some of them to develop their own regional standard, such as the one envisioned by the then Sarpsborg-based high-school teacher Leiv H. Wettre in 1947.<sup>52</sup> That same year, Eng distributed the preliminary translated version to the head of the educational psychology services in Oslo and the psychiatrist Nic Waal.<sup>53</sup> Nevertheless, there was still a considerable way to go at

48 Johs. Sandven, *Intelligensmåling ved gruppeprøver, Avhandlinger fra Universitetets pedagogiske forskningsinstitutt* (Cappelen, 1943).

49 Cecilie Engebriktsen, later Murphy, tested using Eng's material in 1943 and 1944. See Cecilie Engebriktsen, 'Litt om småbarnstesting', in *Fra Pedagogikkens Arbeidsfelt: Skrift til professor dr. Helga Eng på 70-årsdagen 31. Mai 1945* (J. W. Cappelens forlag, 1945), pp. 80–85.

50 See Bærum Skolehjem to Pedagogisk Institutt, Sandvika, 25 April 1951, folder marked 'Korrespondanse til og fra Instituttet i tiden 23. desember 1950–20 desember 1951', PFI UiO.

51 The following participated by gathering data on the specified age groups. Two- and three-year-olds: Else Marie Westh Christensen (known as Edda). Four- and five-year-olds: Cecilie Engebriktsen (later Murphy). Six and seven-year-olds: Elfrida Høygård. Eight- and Nine-year-olds: Turi Sverdrup Lunden. Ten- and eleven-year-olds: Lars Beite. Twelve- and thirteen-year-olds: Else Cordtsen. From fourteen to eighteen, including adult age groups: Per Rand. In addition, Liv Nordgaard, Gunvor Rand and Solveig Ellingsen participated in the testing. All are credited on the frontispiece of the published test, although the specifics of their work were never specified. Some of the students' theses, however, are available from the Library of the University of Oslo.

52 Leiv H. Wettre to Helga Eng, 8 May 1947, in folder marked 'Skriv fra studentene 1945–1948', PFI UiO.

53 Helga Eng to Leiv H. Wettre, 24 May 1947, in folder marked 'Svarskriv til studentene 1946–1948', PFI UiO.



Eng's (delayed) retirement in 1948.<sup>54</sup> The real work on the adaptations only progressed around 1951. The delay was caused by the rather unsystematic manner in which the data was gathered. For example, the results obtained by individual students were not systematically recorded, and the representative sampling of schoolchildren was not uniform.

Johannes Sandven continued the effort to make a Norwegian version of the Terman–Merrill test, but he realized that Eng had never asked Terman or his publishers for any permissions. Embarrassed, he wrote Terman an apology.<sup>55</sup> He was, nevertheless, met with American realities, which implied that the copyright of the tests was not Terman's sole possession, but resided with the publishing company. Terman replied, 'I shall personally be glad if a good Norwegian translation and standardization can be carried out. However, the final decision concerning giving the permission lies with the Houghton Mifflin Company'.<sup>56</sup> The implication of this was that Terman could not decide on the terms for adapting the Terman–Merrill test; it was up to his publisher. By then, Terman was a publishing veteran with ties to all major American textbook and test publishers, be it Houghton Mifflin, World Book Company or McGraw-Hill. As such, he also knew that Houghton Mifflin's response would be to insist on a royalty.

Neither could this have surprised Sandven, who was already locked in negotiations with another American test publisher, the California Test Bureau, regarding the possibility of using parts of their California Test of Mental Maturity and California Capacity Questionnaire in some of his alternative survey work.<sup>57</sup> Sandven also knew there was no reason to believe the tests were economically viable in Norway. He actually thought very few copies of the test would be sold, and needed financial support to get the printing done. However, to Houghton Mifflin, this was completely unprecedented, and they could not agree with Sandven's

54 A 1949 copy of her translation is held by the Oslo University Library.

55 Johs. Sandven to Lewis M. Terman, Oslo, 18 October 1951, in folder marked 'Korrespondanse til og fra Instituttet i tiden 23 desember 1950–20 desember 1951', PFI UiO.

56 Johs. Sandven to Lewis M. Terman, Oslo, 18 October 1951, in folder marked 'Korrespondanse til og fra Instituttet i tiden 23 desember 1950–20 desember 1951', PFI UiO.

57 Ernest W. Ties to Johs Sandven, Hollywood, 30 January 1950, in folder marked 'Skriv vedr instituttet fra 1. juli 1948–22 februar 1950', PFI UiO.

insistence that it should be provided for free. '[W]e cannot believe that any reasonable publisher would object to paying a nominal sum', the American publisher wrote.<sup>58</sup> In a letter to Sandven, the publisher made the reason quite explicit: 'The tests are one of our most valuable properties, and much as we would like to, we simply cannot let them go for use in Norway without royalty.'<sup>59</sup> The proposition was the payment of a nominal royalty: A cash advance of 50 \$ and an agreement to pay a royalty of 6 percent of the list or retail price of the work.

Sandven, however, continued to argue his case, contending that the project would be loss-making, even more so with Houghton Mifflin's demands. He argued that 'if the claim on royalty is maintained, there is, as far as I can see now, scarcely any other procedure possible than to publish a mimeographed edition of fifty or hundred copies to meet the most pressing need'.<sup>60</sup> His publisher, Cappelen, agreed with him:

In this country, whose total population only slightly exceeds that of Boston with its surrounding districts, we cannot regard the task of publishing works like the Stanford–Binet test as anything but a liability and an honourable responsibility rather than a business proposition.<sup>61</sup>

Eventually, Richard N. Clark of Houghton Mifflin reluctantly approved the royalty-free publication. However, his angle was quite explicit: 'We make the exception in the case of the first edition because of the unusual circumstances involved and in the hope that by the time it is exhausted the test will be well enough established in Norway to permit the payment of a royalty on subsequent editions'.<sup>62</sup> To them, the longevity of tests could still make the Norwegian translation into a viable business proposition and a possible source of royalties.

The publisher Cappelen issued the Norwegian translation in 1954. With this,

---

58 Richard N. Clark to Josh Sandven, Boston, 21 April 1953, in folder named 'Korrespondanse til og fra Instituttet I tiden 1 oktober 1952–31 mai 1953', PFI UiO.

59 Richard N. Clark to J. W. Cappelens forlag, Boston, 21 April 1953, copy held in 'Korrespondanse til og fra Instituttet I tiden 1 oktober 1952–31 mai 1953', PFI UiO.

60 Johs. Sandven to Houghton Mifflin, Oslo, 13 April 1953, copy in folder marked 'Korrespondanse til og fra Instituttet i tiden 1 oktober 1952–31 mai 1953', PFI UiO.

61 J. W. Cappelen to Houghton Mifflin Company, Oslo, 29 April 1953, copy in folder marked 'Korrespondanse til og fra Instituttet i tiden 1 oktober 1952–31 mai 1953', PFI UiO.

62 Richard N. Clark to Johs. Sandven, Boston, 27 May 1953, in folder marked 'Korrespondanse til og fra Instituttet i tiden 1. Juni 1953–31. Desember 1953', PFI UiO.

Cappelen became the official publisher of the American revision of a French test, translated and adopted for use in Norway at a public university, subsidized by public funding for the printing. Like the US “original”, additional products were made available by the publisher, such as scorecards and printed cards. However, there was more to the import than striking a deal with an American publisher. It also had to fit Norway linguistically, culturally, and materially.

## Traveling Light

When “moved” to Norway, the 1937 Terman–Merrill test revision traveled light. For example, no one intended to manufacture a Norwegian version of the elaborate box, nor did it contain toy-like miniatures and test material. Furthermore, the test was reduced by half, as only one of the two alternate sets of test items that made up the original version was included in the Norwegian version. Furthermore, the slim Norwegian volume that specified the test carried nowhere near as much information as the American original—it was more a companion to the original volume than a complete translation. The soft binding of the book was of such a dubious quality that nowadays, it is almost impossible to come across a used original that has not been rebound or, at least, clothed.

When a Swedish version was first published in 1943, after being translated, adapted, and standardized by Alice Hellström, the choice was the exact opposite—it was published as a literal translation of Terman and Merrill’s textbook, while information regarding the Swedish translation and statistical adjustments of the test was added as an appendix. The appendix provided tables similar to those found in the American original on statistical standardization. The Swedish volume also mimicked the format of the American original by opting for a hardback binding, a color, and a page size comparable to the original. However, whereas the Norwegian version was published and distributed by a regular publisher, the Swedish version was a private print.<sup>63</sup>

The adaptations in the Norwegian edition were far removed from the technical sophistry that marked the original. All record-keeping and calculations done by the test adaptors were made with paper and pencil; there were no punch cards or Holleriths involved.

---

63 Lewis M. Terman and Maud A. Merrill, *Intelligensmätning: handledning i bruket av de nya omarbetade Stanford-Binet-proven för intelligensundersökning*, trans. by Alice Hellström (Föreningen för sinnesslöa barns vård, 1943).



Fig. 9.2 Material from the standardisation of the Norwegian adaption of the 1937 Terman–Merrill test. Photographer Helge Brekke/Museum of University History (MUV), KHM, University of Oslo. All rights reserved.

As introduced above, the material was worked out by primarily graduate students, covering the tests assigned to the age groups from two to eighteen, with fifty children in each group.<sup>64</sup> They adapted the language of Eng's original translation and carried out testing on what could, at best, be considered a representative sample. The graduate students gathered the frequency of which test items the kids were able to get right. The samples were limited to pupils in Oslo to the extent that the result was admittedly an 'Oslo standard' more than the promised Norwegian standard.<sup>65</sup> The results were submitted individually as part of their studies.

The general impression is that the work could have been more systematic. Even the test developers admitted to this, as noted by Per Rand when he was working out some of the final statistical features of the Norwegian adaptation, such as the reliability and validity scores: 'This could have been a great work *if* only the management and the cooperation had been sorted out from day one', he wrote in a short note

<sup>64</sup> There were exceptions to this. As in the original, from two to four years old, the groups were categorized in steps of six months. For two- and two-and-a-half-year-olds, there were 27 and 44 children tested, respectively. From fifteen to eighteen, there were only 30 children tested. See Sandven, *Norsk standard av Terman–Merrills Stanford-revisjon av Binet-prøvene*, p. 189.

<sup>65</sup> Johs Sandven to Arthur Lillegård, Oslo, 272 October 1956, in folder marked 'Diverse', PFI UiO.

to the project manager Johs Sandven in the summer of 1951. He found his task a rather 'sorry state of affairs'.<sup>66</sup>

There seems to have been next to no coordination among the students, beyond using the Eng translation. For example, different methods of obtaining a representative sample were used for the different age groups, albeit a standard was eventually devised by Lars Beite, one of the participating students. It was left to the students concerned with the trials to devise a method to approximate the sampling method used in the American original, where parents' occupations were used as a proxy for something akin to socioeconomic status. The geographic location of different schools was used as a secondary and relatively intuitive index. Furthermore, the sample sizes varied across the age groups, with the youngest age group having the smallest sample. However, the lack of systematic record-keeping proved to be the most severe weakness of the work—some kept individual records, while others only recorded the total scores for each test item. Although the result achieved some level of consistency, it also had some glaring omissions. For some age groups, the statistical details were transpositions of the American originals; for others, they were based on sums derived from the individual studies.<sup>67</sup>

As decided from the beginning of the project, only one of the two test series (often referred to as scales or forms) was adapted to Norwegian. The American original consisted of one L-form and one M-form, named after Lewis (Terman) and Maud (Merrill), respectively. The L-form was originally more like the 1916 revision. Helga Eng chose to work on the M-form. This made it easier to replace the items that seemed particularly out of place or scored particularly badly in terms of the obtained passing frequencies from the alternative form in the original was tried out. This was done in seven cases out of the 122 items that were included in the Norwegian series. Three items were completely new. The result was that the Norwegian edition was a convergence between the two forms of the American original, albeit only in a few items.

66 Note by Per Rand filed in folder marked 'Korrespondanse til og fra Instituttet i tiden 23. desember 1950–20 desember 1951', PFI UiO.

67 Some remains of the projects are stored in the historical collection of the Department of Education, University of Oslo. This is filed in a box referenced as 'HE.K84.IPED. INST.Test.AB.Pa.Ter.01'. Regarding the lack of details and the consequences for the statistical standardization, this is elucidated in Sandven, *Norsk standard av Terman–Merrills Stanford-revisjon av Binet-prøvene*, p. 189.

When compared to the Swedish adaptation, published in 1943, the Norwegian adaptation was both quite different and actually more systematic. The Swedish adaptation was concerned with the L-form, and the sample size varied even more considerably from age group to age group—twenty-four nine-year-olds to seventy-eight ten-year-olds. In the Swedish version, some of the samples were merely tried out with test items from an older test edition.<sup>68</sup> Many of the considerations made by the Norwegian “adapters” were economical. For example, Johs Sandven formulated it as follows when he communicated with Houghton Mifflin regarding the possibility of manufacturing a Norwegian version of the boxed set of test material: ‘Because of the very restricted market for such material in Norway, we have found it most economical for Norwegian purchasers to use the material already available’.<sup>69</sup> The Norwegian publisher never made an official box. Even the process of purchasing that material was something the Norwegian developers and their publisher were keen to relinquish, as they came to rely on Swedish imports and distribution, as we know from copies stored in the archives of the Department of Education at the University of Oslo. An order had to be placed with the Stockholm-based Skriv & Ritboksaktiebolaget to obtain most of the auxiliary objects. The ordering procedure for the items that should be obtained from Sweden was far from straightforward. When the items were ordered from Sweden, one unlucky purchaser was even challenged with a demand to register as an official importer of toys and an additional fee. Indeed, the publisher would have to strike a special deal with the Norwegian Customs Service and the Ministry of Finance just so those (few) ordering the toy-like material from Sweden would not become actual toy importers in the registers of the customs service.<sup>70</sup>

However, some of the items used in the test had to be purchased individually, such as a flag, a locomotive, and a rotary whisk. All of these could presumably be found at a Norwegian toy store, as the test developers only advertised that ‘if the demand is high, the supply might be serviced by the Department of Education, Oslo’.<sup>71</sup>

68 Terman and Merrill, *Intelligensmätning*.

69 Johs. Sandven to Houghton Mifflin Company, Oslo, 15 May 1953, in folder marked ‘Korrespondanse til og fra Instituttet i tiden 1 oktober 1952—31 mai 1953’, PFI UiO.

70 Skolebokavd. F/BP to Johs Sandven, 31 January 1955, JWC NB.

71 On a printed inlay of the book.



Fig. 9.3 The Norwegian assembly of a Terman-Merrill test. Photographer Helge Brekke/Museum of University History (MUV), KHM University of Oslo. All rights reserved.

Indeed, for a short period, the department at the University of Oslo became purveyors of toy trains and dollhouses. However, their supply had stopped by 1957. As the last items—a model train and a toy shoe—were shipped off, those in need were advised to ‘get the other parts from somewhere else’.<sup>72</sup> Consequently, three years after its publication, the test’s institutional home quite explicitly removed any responsibility for the possibility of obtaining the complete test.

## Tests Wanted

In October 1951, Turi Sverdrup Lunden (later Sverdrup Berg), a former student of Helga Eng, wrote a letter to Johannes Sandven at the University of Oslo. She had just visited a bookstore in Oslo that specialized in trading with educational and professional material. Here, she received news about a forthcoming publication of the Norwegian

<sup>72</sup> From Svanhild Kronkvist to the psychologist at Emma Hjort, 12 April 1957, in folder marked ‘Korrespondanse til og fra Instituttet, 11 september 1956–juni 1957’ PFI UiO.

revision of the Stanford–Binet.<sup>73</sup> She was no stranger to this, as she was one of the graduate students who had tried out Eng’s translation a few years earlier. Indeed, she was one of the collaborators that the project leader had to pester about the lack of adequately kept journals and records before the adaptation could be finalized.<sup>74</sup>

In 1947, she had made extensive tests with children aged eight and nine, complementing the other students engaged in the test trials. She found the work inspiring, although she admitted that ‘arriving at the right atmosphere for the testing was the most interesting part of the challenge and the one I learned most from’.<sup>75</sup>

By 1951, Sverdrup Lunden headed the newly established *Barnevernskolen* in Bergen, which provided preschool teachers with educational training. She also worked at *Barneklubben* [the Children’s clinic at the Haukeland hospital], also in Bergen. While in Bergen, she had not closely followed the final stages of the Terman–Merrill adaptation project.

When visiting Oslo, however, she claimed that the clerk at the bookstore had been informed that the publication was nearing completion—and even more important to her, he had been informed about the possibility of publishing Norwegian record booklets. To Sverdrup Lunden, the record booklet was an essential part of the test. Her work at the clinic relied on importing a Swedish edition of the booklet, as she conducted four to six tests on children a week, and the booklet was an essential component of her testing system. She also considered the cost of the Swedish booklets to be expensive and cumbersome to obtain. Furthermore, the Norwegian edition of the test would not “line up” according to the Swedish record booklet, as the sequence and the wording of the test items could be different. About the record booklet, she wrote to Sandven in Oslo that she was ‘very interested’.<sup>76</sup> Sverdrup

73 Turi Sverdrup Lunden to Johs. Sandven, Bergen, 4 October 1951, in folder marked ‘Korrespondanse til og fra Instituttet i tiden 23. desember 1950–20. desember 1951’, PFI UiO.

74 This was acknowledged in Sandven, *Norsk standard av Terman–Merrills Stanford-revisjon av Binet-prøvene*, p. 189.

75 Turi Sverdrup Lunden, ‘Intelligensmåling med individualprøver: måling av 8- og 9-åringer i Oslo med reviderte Stanford-Binet 1937’ (Universitetet, 1949), p. 30.

76 Turi Sverdrup Lunden to Johs. Sandven, Bergen 4 October 1951, in folder marked ‘Korrespondanse til og fra Instituttet i tiden 23. desember 1950–20. desember 1951’, PFI UiO.



Lunden hoped that the coordinated effort in Oslo would publish the standardized instructions and results, the auxiliary material needed for the tests, and the record booklet.

Sverdrup Lunden's need for the record booklet reveals how the test was used. In many ways, the booklet would be easy to make up yourself. It was eleven pages, where all the items were given a form. It could be made as a stencilled copy (even though this was before the invention of the Xerox machine), drawn up with pen and paper, or forged with a lithograph—or even using see-through paper. However, the usability and formality of the official booklet fitted the increasingly formalized use of the test. Indeed, it somehow facilitated the regimented use of the test by “filling it in” as a file, with form fields for all official information printed on the front of the booklet. The blank spaces on the back of the booklet were also regularly used to store more informal information.<sup>77</sup> In many ways, the booklet made the test more like ‘a material, procedural system’, as in Peter Galison's interpretation of the history of the Rorschach test. It indicated how to organize the testing session. It directed and dictated.<sup>78</sup>

When Sverdrup Lunden wrote the letter, no plans had been made to make a Norwegian record booklet.<sup>79</sup> Upon final publication, however, a record booklet was among the auxiliary products made available from the publisher. In the same letter, Sverdrup Lunden revealed more about her work at the clinic and, by doing this, more about the use of the test. Her appointment and testing practice were all temporary, ‘while we await the announcement of the full and permanent position in psychology. We also await the announcement of a position as school psychologist here [in Bergen],’ she wrote.<sup>80</sup> Psychologists were considered the experts in testing. But whether Sverdrup Lunden could be regarded as one was not all that clear at the time. Her background was in educational research, although it was heavily steeped in psychology. She did not know whether she was qualified for the upcoming position.

77 A set of examples has been found in a box referenced as ‘HE.K84.IPED.INST.Test. AB.Pa.Ter.. 01’, IPED UiO.

78 Peter Galison.

79 U.n. to Turi Sverdrup Lunden, copy of letter, Oslo 8 October 1951, in folder marked ‘Korrespondanse til og fra Instituttet i tiden 23. desember 1950–20. desember 1951’, PFI UiO.

80 Turi Sverdrup Lunden to Johs. Sandven, Bergen, 4 October 1951, in folder marked ‘Korrespondanse til og fra Instituttet i tiden 23. desember 1950–20. desember 1951’, PFI UiO.

The letter also reveals something about timing—neither the psychologists nor the positions were “in place”. The 1940s and 1950s were a period of significant transformation in psychology and testing in Norway.<sup>81</sup> The organization of the social and educational system, both at the municipal and national levels, changed considerably during this period. The establishment of educational psychological services was first shaped by local contexts, particularly in larger cities like Oslo, Trondheim, and Bergen, with considerable differences among them in terms of services provided and the kind of expertise employed. As such, the test would have to fit into a system that was not yet fully established or institutionalized.

Sverdrup Lunden also revealed doubts and worries about her practice as an intelligence tester at a child clinic, a distinctively different institutional setup from the nascent educational psychological offices. She wrote:

It is quite interesting to work with the children at the clinic, as I have limited training in this field. At the same time, it is unsatisfactory because it can never be fully adequate. Working together with physicians is not always straightforward, as they ask for IQ even when the child cannot speak or move at all. If they receive an IQ, they are placing a label on the child, who will live with it for the rest of their lives, proclaiming a result that, incidentally, was determined despite all possible errors.

As evident in Sverdrup Lunden’s doubts and worries, the use of intelligence testing in the early 1950s can be understood as a conflict between professional expertise, particularly that of medical, educational, and psychological experts. Such an interpretation also holds a prominent place in the established Norwegian historiography regarding intelligence testing, particularly in the work of the historian Eva Simonsen.<sup>82</sup> Here, intelligence tests often animate the distinction between medical and educational expertise, and the discussion of who should be given the means (and the permission) to use intelligence tests. The test adaptation of the Terman–Merrill revision originated in a

---

81 Jan Froestad and Bodil Ravneberg, ‘Education Policy, the Norwegian Unitary School and the Social Construction of Disability’, *Scandinavian Journal of History*, 31.2 (2006), 119–43.

82 Eva Simonsen, *Vitenskap og profesjonskamp: opplæring av døde og åndssvake i Norge 1881–1963* (Unipub forlag, 2000).

department concerned with educational applications of the test, but as indicated in the above, its practical application was far from limited to the educational field. Moreover, potential for conflict was never far away when it came to questions of who was best suited to use such tests.

In the Department of Education at the University of Oslo, for instance, the conflicts between general teachers and those specialized in psychology or educational psychology came to shape the final distribution of this test. In the department archives, several letters ask how to obtain the Norwegian version of the Terman–Merrill test. The formulaic answer typically provided by the department pointed to conditions in which only those with either educational, psychological, or psychiatric training were allowed to order the test from the publisher.<sup>83</sup> Teachers, in general, were not considered qualified—a stand strikingly different from that held by the test originator Lewis Terman, who made the test available to ‘the rank and file of teachers, physicians and social workers’.<sup>84</sup> The correspondence of the Department of Education includes several rejections of approval, even in cases where the applicant, a deputy head teacher, argued that there was no way a certified school psychologist would venture to such far-away places. Such was the determination to keep the test out of the hands of regular teachers that Sandven instructed the publisher to reject the sale from all large, specialized booksellers, even those who were already selling the American original.<sup>85</sup> Not even libraries were allowed to carry the test.<sup>86</sup>

It took some convincing to get the publisher to agree to this rather strict sales policy. Indeed, the policy would, in reality, make the test part of a mail-order service for a publisher accustomed to wholesalers and booksellers. This was in contrast to group tests, which were considered appropriate for general teaching training. For example, the published

---

83 Johs. Sandven to Olena Bö, Oslo 13 August 1956, in folder marked ‘Korrespondanse til og fra Instituttet (1954–57), A-K’, PFI UiO.

84 Lewis M. Terman, *The Measurement of Intelligence: An Explanation of and a Complete Guide for the Use of the Stanford Revision and Extension of the Binet–Simon Intelligence Scale*, Riverside Textbooks in Education (Houghton Mifflin Company, 1916).

85 Hans Christian Welleø to Universitetets Pedagogiske Forskningsinstitutt, Oslo, 10 June 1955, in folder marked ‘Korrespondanse til og fra Instituttet i tiden juli 1953–1956, L–Ø’, PFI UiO.

86 Johs. Sandven to Johanne Martens, Universitetsbiblioteket i Bergem Oslo 23 October 1956, in folder marked ‘Korrespondanse til og fra Instituttet, 11 september 1956–juni 1967’, PFI UiO.

version of Kuhlman–Anderson was free to purchase at a bookstore. The secrecy of the Norwegian version of the Terman–Merrill test was portrayed as a virtue—in an advertisement of the test it was argued that it was only by restricting its circulation to experts that the test could remain valuable.<sup>87</sup> Without the strict limits placed on the sales, the test would be made available for all to know about and, presumably, to learn to use.

Sverdrup Lunden was not alone in awaiting the Terman–Merrill. Typically, psychologists and teachers regularly asked when they could get their hands on it, and complained about what was available to them.<sup>88</sup> Questions about the lack of suitable tests for Norwegian children were raised in the press. An expression of this was when the first educational psychologist in the city of Trondheim was interviewed by the *Adresseavisen* daily: ‘There are few suitable tests around today’, she argued, adding that at least there was hope, as the Terman–Merrill was in the works in Oslo.<sup>89</sup>

## Using the Test

When the test was published in the autumn of 1954, it became widely used, despite the restrictive sales model described above. The institutional context that Sverdrup Lunden was alluding to just three years earlier, had also begun to change. Several more educational and psychological services were established in cities and larger municipalities. For example, in Trondheim, Norway’s third-largest city, the educational and psychological counseling service was established in 1951. By 1955, it consisted of two psychologists and a secretary. By the 1955–56 school year, testing had become a considerable time commitment—458 tests of intelligence and other cognitive abilities were administered, and more than fifty percent of those referred to the counseling service involved testing. Two tests were used: the Norwegian version of the Terman–

87 Stated in various advertisements printed in the journal *Vår skole*, for example in number 42, 16. October 1954.

88 Kjell Never, Vestfold institutt for evneprøving og yrkesrettledning, to Johs. Sandven, Sandefjord 16 January 1952, folder marked ‘Korrespondanse til og fra instituttet, 1 januar 1952–30 september 1952’, PFI, UiO.

89 From an interview with Berit Christensen, the first educational psychologist in Trondheim, interviewed in *Adresseavisen*, 4 July 1951.

Merrill revision and one by Grace Arthur.<sup>90</sup>

This gradual expansion of educational psychological counseling services at the municipal level coincided with a gradual transformation of both public institutions and educational facilities for individuals with intellectual disabilities.<sup>91</sup> At the outset of the 1950s, extensive law reforms established a system of state-run and large institutions for those deemed severely intellectually disabled, and special schools for those deemed '*evneveike*'.<sup>92</sup> This term, which came into use in the postwar period, literally means 'of weak abilities', and held implications close to the American term of 'mental retardation'.<sup>93</sup> The system of large state-run institutions and special schools replaced the interwar system of few formalized institutions for care, but with schools attending to, on the one hand, the 'feeble-minded' and, on the other, '*særskolen*' [special school] for a rather undifferentiated category of pupils. In the 1950s, scores on intelligence tests were one of the referral criteria for segregation; the local counseling service made the referrals. However, the test did not provide hard cut-off points, and the differentiation between references to special schools and care institutions was not made solely based on intelligence test scores.<sup>94</sup>

However, the test would take on a more diverse range of uses. Sverdrup Lunden's unassuming letter pointed towards its use by hospitals, even in cases far removed from the referrals made by the counseling services. One of the most striking repurposing of intelligence testing was commented on by the first psychologist given the opportunity to

---

90 Details about the testing made in Trondheim is found in the annual report of Skolepsykologisk kontor, Trondheim, copy held in a folder marked 'diverse', PFI, UiO. While the name of the Grace Arthur test is not provided in the annual report, it was presumably the Point Scale of Performance Scale that was used.

91 Froestad and Ravneberg; Bjørn Hamre, Thom Axelsson, and Kari Ludvigsen, 'Psychiatry in the Sorting of Schoolchildren in Scandinavia 1920–1950: IQ Testing, Child Guidance Clinics, and Hospitalisation', *Paedagogica Historica*, 55.3 (2019), 391–415. See also the various contributions in *Spesialpedagogikk mellom profesjon og disiplin: En utviklingslinje i tre deler*, ed. by Elisabeth Hesjedal and Bjarte Furnes (Fagbokforlaget, 2023).

92 Tora Korsvold, *Barns verdi: en barndom som evneveik på 1950-tallet* (Abstrakt forlag, 2006).

93 Eva Simonsen, "'Folkeskolens ytterste forpost' - en annen historie om den norske folkeskolen 1945–1970', *I særklasse. Inklusion og eksklusion i grundskolen* (Uddannelseshistorie: Årbok for Selskabet for Skole- og Uddannelseshistorie, 2012), pp. 118–44.

94 See chapter one for details on the interwar practices.

practice at a neurological section in a hospital. Hallgrim Kløve would be tasked with intelligence testing during his days as a psychology student. However, here, the testing was related to identifying localized brain damage. He continued his practice after graduating, only to forge a career as a pioneering neuropsychologist in the US, and later in Bergen in the mid-1950s. However, the Terman–Merrill test would be supplanted by more specialized tools in a neurological context, not least of which is diagnostic imaging.<sup>95</sup> In the 1950s and 1960s, the Norwegian edition was still used in studies of cerebral palsy, primarily because it was the only formally adapted intelligence test available at the time, rather than because it was particularly well-suited for the task.<sup>96</sup>

When considering the use of this specific Norwegian adaptation, it is tempting to trace the intersection between those involved in making the Norwegian edition and the use of the test. Johannes Sandven, for example, would continue a strong interest in quantitative methods and educational assessments as a professor at the University of Oslo. However, he was never involved in further research regarding this specific test. The same applies to Per Rand, who later became a professor of education with a clear quantitative bent, as well. For others, such as Cecilie Murphy, Lars Beite, Elsa Cordtsen, and Else Marie Westh Christensen, the test could have been part of their daily work as teachers, administrators, and educators. Elsa Cordtsen can even be found as a defender of test practices in schools during the so-called ‘IQ controversy of 1959’.<sup>97</sup>

Again, Turi Sverdrup Lunden provides some further clues—from 1951, she headed the educational training for preschool teachers in Bergen. When she described the training, she admitted to the influence of developmental psychology in the training of preschool teachers. However, intelligence testing in general and the Terman–Merrill test in particular seem of little influence here.<sup>98</sup>

Despite the lack of any forceful institutional or personal backing,

95 Interview with Hallgrim Kløve in *Tidsskrift for Norsk Psykologforening*, 1 February 2004, vol. 41, nr. 2, 110–115.

96 Turid Vogt Grinde, *Barn med hjerneskader og vanførhet: livssituasjon og personlighetstrekk*, Norbok, Nordisk psykologis monografiserie 25 (Akademisk forlag, 1972).

97 “Psykologisk undersøkning av born i skulen”, Monday 13 April 1959, NRK Radio. On the “IQ controversy of 1959”, see Thue, chapter 11, this volume.

98 Turi Sverdrup Berg, *Førskolelærerutdanning i Bergen gjennom 40 år*, Bergen Lærerhøgskoles skriftserie (Bergen lærerhøgskole, 1988).

in 1960, the Terman–Merrill revision was considered ‘without a doubt the most used intelligence test in Norway’.<sup>99</sup> However, the widespread use of the test was perhaps more due to the lack of alternatives than to the test itself. In one of the few reviews of the test, however, it was concluded that ‘the Norwegian standard of the Terman–Merrill tests of 1954 does have serious methodological flaws, and that a revision and standardisation is absolutely necessary’.<sup>100</sup> At the same time, after an informal survey among the members of *Norsk Skolepsykologers Forening* [the Norwegian Association of School Psychologists], it was concluded that there was ‘a strong need for an improved verbal test’. Research also indicated a weak correlation between the scores obtained with the Norwegian edition and other less verbally oriented tests.<sup>101</sup> Nonetheless, the test stuck. It was kept on the shelves.

## Conclusions

In 1955, a year after the publication of the Norwegian Stanford–Binet, the second WHO-endorsed revision of the International Classification of Diseases (ICD) was agreed upon. Here, the same Houghton-Mifflin-owned product became part of a definition of “mental retardation”, as cutoff points scored on this specific test were explicated in the manual when grading the categories of idiocy, imbecility, and moron. It was the only test named as a criterion for classifying mental retardation. Not only that, but perhaps more astonishingly, it was the sole criterion listed in the complete manual for anything, as the rest of the manual was little more than a list of agreed-upon names. Besides the proximity of the dates between the Norwegian publication of the revised Stanford–Binet in 1937 and the WHO list, this is a somewhat arbitrary juxtaposition. The ICD was far from being the infrastructure of categories and standards it is today.<sup>102</sup> Furthermore, it was decades away from having any real impact

99 Eiliv Solum and Hans Østby, *Erfaringer med norsk standard av Terman–Merrill–Stanfordrevisjon av Binetprøvene 1954* (1961).

100 Solum and Østby.

101 Margrete Landmark and Turid Grinde, ‘Undersøkelse av en gruppe normale barn med forskjellige intelligensprøver’, *Nordisk Psykologi*, 14.4 (1962), 171–85.

102 The standard work on the ICD is Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences*, Inside Technology (MIT Press, 1999). While there exists a large literature on the history of medical categorization, the observation about the formalization of the “mental retardation” criteria by a specified

on Norwegian practice. Nevertheless, the proximity of the dates reveals the short distance between “the formal” and “the commercial”, between rules and products, between authorized criteria and private goods, and between commodity and law. They can, at times, be the same. Along the way, both the formal and the commercial can be changed—they can be translated, adapted, and standardized anew; they can even be cut in half and significantly reordered.

The malleability so evident when adapting the Terman–Merrill test for use in Norway, in terms of statistical policy, test items, and packaging, testifies to a significant feature of the longevity of intelligence tests. To stay on the shelf, alterations shaped and reshaped the short distance between commodity and law. Despite the lack of any sort of forceful backing, the procedural and material system of the test would continue to provide an essential part of an institutional system. Here, publishers mediated the longevity more than the test designers or those involved in its adaptation. It was the Norwegian publisher that kept the record forms in print, and it was the American publishers that eventually allowed the printing of the Norwegian version to be an official version. However, it was public funding that even made publishing in Norway possible, and by doing so, transformed a French original—revised by Stanford faculty psychologists and published by a large American publishing company—into a modest Norwegian adaptation available to several users. However, the institutional system in which the Norwegian adaptation was taking part was in flux and changing when the Norwegian Stanford–Binet was issued. Here, the sheer availability of the Terman–Merrill test was perhaps the most critical factor, more so than its design and capabilities. However, one could still be tempted to argue that the flexibility that so readily comes to the fore when looking at how it was adapted to a Norwegian context was what gave the test longevity. It was mutable, which is why it was successful. Despite its apparent flaws, it was being used readily.

---

test is, to my knowledge, not commented on in the historical literature.



## Archives

The archival holdings of the Department of Education at the University of Oslo have provided the primary sources for this chapter. This is organized in two distinct sets: A substantial set of correspondence of the Pedagogisk forskningsinstitutt (PFI) was found during the research for this chapter, containing material from 1938 to 1958. As this set held numerous administrative files, it now converges with the general administrative archival holdings of the University of Oslo. In the above, I refer to the original folder names of the correspondence and the following abbreviations: PFI, UiO. In addition, the Department of Education has an organized historical archive, holding both files and historical objects. I reference the following archival identifier, followed by IPED, UiO. Correspondence has also been obtained from the personal papers of Helga Eng, held by the National Library of Norway, Oslo. In the above, this is referenced by HE NB. The correspondence of the publisher J. W. Cappelén is available at the National Library of Norway, Oslo. In the above, I reference this as JWC NB. The Houghton Mifflin Company records, 1866–1968, are held by the Harvard Library (MS Am 2030-2030.4), which includes copyright statements and contract records. From the inventory, however, no material relevant to this chapter is available in these records.

## Bibliography

- Berg, Turi Sverdrup, *Førskolelærerutdanning i Bergen gjennom 40 år*, Bergen Lærerhøgskoles skriftserie (Bergen: Bergen lærerhøgskole, 1988).
- Bergstrom-Katz, Alexandra, 'On Intelligence Tests: Psychological Objects and Their Subjects' (Unpublished PhD, Department of Psychosocial Studies, Birkbeck, University of London, 2023).
- Binet, Alfred, and Théodore Simon, *The Development of Intelligence in Children*, trans. by Elizabeth S. Kite (Baltimore: Williams & Wilkins company, 1916).
- Boake, Corwin, 'From the Binet-Simon to the Wechsler-Bellevue: Tracing the History of Intelligence Testing', *Journal of Clinical and Experimental Neuropsychology*, 24.3 (2002), 383–405, <https://doi.org/10.1076/jcen.24.3.383.981>

- Bowker, Geoffrey C., and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences*, Inside Technology (Cambridge, Massachusetts: MIT Press, 1999), <https://doi.org/10.7551/mitpress/6352.001.0001>
- Brown, JoAnne, *The Definition of a Profession: The Authority of Metaphor in the History of Intelligence Testing, 1890–1930* (Princeton, New Jersey: Princeton University Press, 1992).
- Buros, Oscar K., 'Fifty Years in Testing: Some Reminiscences, Criticisms, and Suggestions', *Educational Researcher*, 6.7 (1977), 9–15.
- Carson, John, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750–1940* (Princeton, New Jersey: Princeton University Press, 2007), <https://doi.org/10.1515/9780691187679>
- Clausen, Johs., 'Intelligensmåling i grupper', *Norsk Pedagogisk Tidsskrift*, 25 (1941), 65–82.
- Darnton, Robert, "'What Is the History of Books?' Revisited', *Modern Intellectual History*, 4.3 (2007), 495–508, <https://doi.org/10.1017/s1479244307001370>
- Emberland, Isak Lønne, 'Den omhuggelig planlagte livsbane: Psykoteknikken og profesjonaliseringen av norske psykologer, 1920–1947' (Unpublished PhD, Oslo Met - storbyuniversitetet, 2024).
- Eng, Helga, *Norsk psykoteknikk i ti år 1925–1935: historisk oversikt* (1935).
- , *The Psychology of Children's Drawing: From the First Stroke to the Coloured Drawing* (London: Kegan Paul and Co., 1931).
- Engebretsen, Cecilie, 'Litt om småbarnstesting', in *Fra pedagogikkens arbeidsfelt: Skrift til professor dr. Helga Eng på 70-årsdagen 31. Mai 1945* (Oslo: J. W. Cappelens forlag, 1945), pp. 80–85.
- Froestad, Jan, and Bodil Ravneberg, 'Education Policy, the Norwegian Unitary School and the Social Construction of Disability', *Scandinavian Journal of History*, 31.2 (2006), 119–43, <https://doi.org/10.1080/03468750600604119>
- Galison, Peter, 'Image of Self', in *Things That Talk: Object Lessons from Art and Science*, ed. by Lorraine Daston (New York: Zone Books, 2004), pp. 257–94.
- Grinde, Turid Vogt, *Barn med hjerneskader og vanførhet: livssituasjon og personlighetstrekk*, Nordisk psykologis monografiserie 25 (København: Akademisk forlag, 1972).
- Hamre, Bjørn, Thom Axelsson, and Kari Ludvigsen, 'Psychiatry in the Sorting of Schoolchildren in Scandinavia 1920–1950: IQ Testing, Child Guidance Clinics, and Hospitalisation', *Paedagogica Historica*, 55.3 (2019), 391–415.
- Helsvig, Kim G., *Pedagogikkens grenser: kampen om norsk pedagogikk ved Pedagogisk forskningsinstitutt 1938–1980* (Oslo: Abstrakt forlag, 2005).
- Hesjedal, Elisabeth, and Bjarte Furnes, eds., *Spesialpedagogikk mellom profesjon og disiplin: En utviklingslinje i tre deler* (Bergen: Fagbokforlaget, 2023).

- Korsvold, Tora, *Barns verdi: en barndom som evneveik på 1950-tallet* (Oslo: Abstrakt forlag, 2006).
- Landmark, Margrete, and Turid Grinde, 'Undersøkelse av en gruppe normale barn med forskjellige intelligensprøver', *Nordisk Psykologi*, 14.4 (1962), 171–85.
- Larsen, Lina Bryn, 'med maaling følger viden': *Intelligensmåling i den norske folkeskolen, 1900–1945*, Hovedfagsoppgave i historie (Universitetet i Bergen, 2006).
- Lofthus, Johan, Augusta Rasmussen, and Bernhard Ribsskog, *Evnemålinger i folkeskolens 1. klasse: Dearborns gruppeprøve* (Oslo: Gyldendal, 1937).
- Lønnå, Elisabeth, *Helga Eng: psykolog og pedagog i barnets århundre* (Bergen: Fagbokforlaget, 2002).
- Lunden, Turi Sverdrup, 'Intelligensmåling med individualprøver: måling av 8- og 9-åringer i Oslo med reviderte Stanford–Binet 1937' (Universitetet i Oslo, 1949).
- McNemar, Quinn, *The Revision of the Stanford-Binet Scale: An Analysis of the Standardization Data* (Boston: Houghton Mifflin, 1942).
- Minton, Henry L, Lewis M. Terman: *Pioneer in Psychological Testing* (New York: New York University Press, 1988).
- Mülberger, Annette, 'Biographies of a Scientific Subject: The Intelligence Test', in *Oxford Research Encyclopedia of Psychology* (Oxford: Oxford University Press, 2020), <https://doi.org/10.1093/acrefore/9780190236557.013.694>
- , 'The Need for Contextual Approaches to the History of Mental Testing', *History of Psychology*, 17.3 (2014), 177–86, <https://doi.org/10.1037/a0037487>
- Østlyngen, Emil, *Aktuelle tester i Norge* (Oslo: Norsk Psykologforening, 1957).
- , *Bibliografi over psykologisk faglitteratur av norske forfattere* (Oslo: Norsk psykologforening, 1942).
- Pedersen, Marie, *Intelligensprøving av barn: Dearborns gruppetest* (Oslo: Gyldendal, 1933).
- Raven, James, *Publishing Business in Eighteenth-Century England* (Martlesham, Suffolk: The Boydell Press, 2014), <https://doi.org/10.1515/9781782043720>
- Ribsskog, Bernhard, *Evneprøve for 1. klasse i folkeskolen: standardisert på grunnlag av Kuhlmann-Andersons gruppeprøve* (Oslo: Gyldendal, 1941).
- Richardson, Ken, *Understanding Intelligence*, Understanding Life Series (Cambridge: Cambridge University Press, 2022), <https://doi.org/10.1017/9781108937757>
- Sandven, Johs., *Intelligensmåling ved gruppeprøver*, Avhandlinger fra Universitetets pedagogiske forskningsinstitutt (Oslo: Cappelen, 1943).

- , *Norsk standard av Terman–Merrills Stanford-revisjon av Binet-prøvene* (Oslo: Cappelen, 1954).
- Simonsen, Eva, “‘Folkeskolens ytterste forpost’ - en annen historie om den norske folkeskolen 1945 - 1970’, *I særklasse. Inklusion og eksklusjon i grundskolen* (Uddannelseshistorie: Årbok for Selskabet for Skole- og Uddannelseshistorie, 2012), pp. 118–44.
- , *Vitenskap og profesjonskamp: opplæring av døde og åndssvake i Norge 1881–1963* (Oslo: Unipub forlag, 2000).
- Solum, Eiliv, and Hans Østby, *Erfaringer med norsk standard av Terman–Merrill-Stanfordrevisjon av Binetprøvene 1954* (1961).
- Terman, Lewis M., *The Intelligence of Schoolchildren: How Children Differ in Ability. The Use of Mental Tests in School Grading and the Proper Education of Exceptional Children* (Boston: Houghton Mifflin Company, 1919).
- , *The Measurement of Intelligence: An Explanation of and a Complete Guide for the Use of the Stanford Revision and Extension of the Binet-Simon Intelligence Scale*, Riverside Textbooks in Education (Boston: Houghton Mifflin Company, 1916).
- , *The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence* (Baltimore: Warwick & York, Inc., 1917).
- Terman, Lewis M., and Maud A. Merrill, *Intelligensmätning: handledning i bruket av de nya omarbetade Stanford-Binet-proven för intelligensundersökning*, trans. by Alice Hellström (Stockholm: Föreningen för sinnesslöa barns vård, 1943).
- , *Measuring Intelligence: A Guide to the Administration of the New Revised Stanford-Binet Tests of Intelligence* (Boston: Houghton Mifflin Company, 1937).
- , *Stanford-Binet Intelligence Scale; Manual for the Third Revision Form L–M* (Boston: Houghton Mifflin Company, 1960).
- Trent, James, *Inventing the Feeble Mind: A History of Intellectual Disability in the United States* (Oxford: Oxford University Press, 2016), <https://doi.org/10.1093/med/9780199396184.001.0001>
- Vismann, Cornelia, *Files: Law and Media Technology*, Meridian (Stanford: Stanford University Press, 2008).
- Young, Jacy L., ‘Test or Toy? Materiality and the Measurement of Infant Intelligence.’, *History of Psychology*, 18.2 (2015), <https://doi.org/10.1037/a0038862>
- Zenderland, Leila, *Measuring Minds: Henry Herbert Goddard and the Origins of American Intelligence Testing* (Cambridge: Cambridge University Press, 1998).



## 10. “Culture” and Representability in the Norwegian Standardization of WISC-R

*Jon Røyne Kyllingstad*

---

A 1998 study showed that children with immigrant backgrounds were highly overrepresented among pupils that received special education measures in Oslo schools.<sup>1</sup> Some years later, three studies suggested that this was related to IQ testing. Minority children with learning difficulties were assessed with IQ tests intended for majority children, and test results were often interpreted without considering the background of the children. In some cases, children who needed training in the Norwegian language or were entitled to teaching in their own mother tongue, instead were diagnosed as having learning difficulties or intellectual disabilities.

Most of these children had been tested with one specific test, namely the WISC-R (Wechsler Intelligence Scale for Children—Revised). This test was launched in the USA in 1974 and appeared in a Norwegian version in 1978. It was the second of five revised versions of the WISC test that have been published from 1949 to the present. These tests are probably the world’s most widely used ability tests for children. In Norway, they are often used by the *Pedagogisk-psykologisk tjeneste* [Educational-Psychological Services, EPS] to assess children with special educational needs. The WISC-R was the first WISC version that was officially translated, adapted, and standardized for use in Norway, and it is also, to date, the most thoroughly adapted Norwegian version of a WISC test. This chapter examines how the originally American WISC-R was transformed into a Norwegian one.

---

1 Thomas Nordahl and Terje Overland, *Idealer og realiteter: Evaluering av spesialundervisningen i Oslo Kommune* (Oslo: NOVA Rapport 20, 1998).

A Wechsler scale is a relative scale. An individual's IQ score indicates how much their test results deviate from the average of the population to which the individual belongs. WISC-R was originally developed as a tool to compare the test results of an individual school child in the USA with the average test results among US schoolchildren. To transform the test into a Norwegian one, the test instructions and test items had to be translated and culturally adapted so that the tasks would be comprehensible within a Norwegian school child's linguistic and cultural frame of reference. The test also had to be 'normed', which means that the calculation procedures for converting raw test scores into an IQ score were adjusted, based on the average distribution of test scores among Norwegian school children. However, since it was impossible to try out the test on all Norwegian schoolchildren, both the cultural adaptation and the norming of the scale were based on a sample of schoolchildren who were to represent all Norwegian school children.

These operations are normal procedures in what is often referred to as the standardization of a test. But they raise some interesting questions regarding two topics that will be discussed in this chapter, namely representation and the notion of culture. What were the criteria for representativeness that lay behind the selection of schoolchildren? Why were these criteria chosen, and what can this tell us about the societal, institutional, practical, and political contexts—and the preconceived notions about Norwegian society and culture—which formed the conditions for the standardization process? Furthermore, which notions of "culture" was the cultural adaptation of the test based on? The question is not primarily how the test developers themselves explicitly defined the concept of culture, but rather how concepts of culture and cultural differences were operationalized through the actual work of cultural adaptation.

## Debates About Cultural Adaptation and Fairness of Tests

Before going into the Norwegian adaptations, let us first look at the historical context of the original American version of WISC-R as a way to further clarify the problem of cultural bias in intelligence testing.

The breakthrough of IQ testing in the USA around World War One had been characterized by strong hereditarianism, a connection to

eugenics, and a tendency to explain social and racial inequality as a product of innate and stable mental capacities (see, for instance, Porter, chapter 13, this volume). Although these ideas were never uncontested, it was mainly the 1930s and the postwar years that saw a systematic and eventually dominating critique against hereditarianism and racial theories. Environmental explanations of intelligence came to the fore, stressing that an individual's IQ score was the product of both nature and nurture—and, importantly, that average IQ differences between social groups (such as "races") was the product of social, economic, and educational disparities, not inborn racial differences. Around 1969, however, a heated controversy about race and IQ erupted again, sparked by a paper by the educational psychologist Arthur Jensen. He proposed that environmental factors could only explain parts of the difference in average IQ scores between races and social classes, and that educational programmes intended to increase the IQ scores of disadvantaged groups were doomed to fail.<sup>2</sup>

We will not go into details about this well-known controversy. But in order to identify some questions that are relevant to the Norwegian adaptation project, we will look at some of the viewpoints put forward by key participants in the debate, including individuals involved in the 1974 revision of the WISC test. David Wechsler, who invented the Wechsler tests, Alexander G. Wesman, who was the test division director of The Psychological Corporation that published the test, and Alan S. Kaufman, who played a key role in the WISC-R revision, advocated what Porter describes as an interactionist view—namely that intelligence, as measured by an IQ test, is a product of both 'native endowment' and learning, and that learning is influenced by socio-economic and cultural factors. This insight was a core aspect of 'intelligent testing', a slogan used by both Wesman and Kaufman to describe how IQ test scores should be interpreted.<sup>3</sup> A key point was that intelligence is an attribute, not an entity, and that IQ tests measure both inborn aptitude and learned

---

2 See for example, Raymond E. Fancher, *The Intelligence Men: Makers of the IQ Controversy* (New York: WW Norton & Co., 1985).

3 The slogan was first used by Wesman in a 1967 presidential address to the American Psychological Association's Section for Measurement, Assessment, Statistics, and Testing. Wesman was also a member of a task force of the American Psychological Association on testing of minorities and underprivileged groups. Alan S. Kaufman, *Intelligent Testing with the WISC-III* (New York: Wiley, 1979).



achievement, but that the test by itself cannot discriminate between the two types of causalities.<sup>4</sup>

In a 1969 presidential address to the American Psychological Association's Section for Measurement, Assessment, Statistics and Testing, Wesman dismissed the idea that it is possible to create 'culture-free' or 'culture-fair' tests. Attempts to create such tests were based on 'the tacit assumption that "native intelligence" lies buried in pure form deep in the individual', and can be 'uncovered' through 'ingenious mining'. However, this assumption was baseless, since 'intelligence comprises learning experiences' that are embedded in culture.<sup>5</sup> The same point was made in Kaufman's book *Intelligent Testing with the WISC-R* (1979), where WISC-R is described as a 'culture loaded' achievement test, since it 'measures what the individual has learned', and 'since learning occurs within a culture'. Kaufman's book became highly influential as a key to the interpretation of WISC-R test scores, as we will see, in Norway.<sup>6</sup>

Therefore, both Wesman and Kaufman acknowledged that IQ tests are culturally loaded, and that individuals should only be assessed with a test developed for their own group. This raises questions about the definition and delineation of the group to which the test should be normed and to whose culture it should be adapted. These are methodological questions, but they are steeped in normative and political issues about culture, race, and social class—issues that were hotly debated at the time when the WISC-R saw the light of day.

African American anti-racist activists, such as the psychologist and founder of the Black Psychologists Association, Robert L. Williams, strongly argued, for example, that WISC and other standard IQ tests were 'culture specific tests', adapted to 'white middle class' culture. In the early 1970s, he put forward an alternative test adapted to African American culture. The Black Intelligence Test of Cultural Homogeneity (BITCH-100) was a verbal test, meant to determine the test taker's ability to function symbolically in their own cultural environment. It was adapted to the language and culture of black children from low- and middle-income groups in the city of St. Louis, and it produced test

---

4 Alexander G. Wesman, 'Intelligent testing', *The American psychologist*, 23.4 (1968), 267–274 (pp. 268–69), and Kaufman, *Intelligent Testing with the WISC-III*.

5 Wesman, 'Intelligent testing', p. 269.

6 Kaufman, *Intelligent Testing*, pp. 11–12.

results that were in sharp contrast to traditional IQ testing—the average African American child scored significantly above the average white child.

Williams was in line with Wesman when arguing that 'culture specific tests' could determine the child's ability to function symbolically in 'his own culture and environment'. Williams claimed, however, that a culture-specific test, by measuring a child's capability for conceptual thinking in their own "culture", could also predict how well that child would 'master those aspects of the elementary school curriculum, requiring this dimension of ability'.<sup>7</sup> Wesman, in contrast, held that a culture-specific test adapted to a subculture could only measure how much of a specific subculture a person had absorbed. It could not determine a child's potential for future learning of the content of what he described as 'the more general culture', and thus to succeed in school.<sup>8</sup>

In 1978, sociologist Jane Mercer proposed another solution to the problem of cultural test bias—the SOMPA, or System of Multicultural Pluralistic Assessment. It was meant to combine medical, sociological, and pluralistic perspectives to assess schoolchildren's need for special education. A key principle was that a child should only be compared with other children from the same sociocultural background. The SOMPA battery included a standard WISC-R test, which, according to Mercer, measured partly the child's potential for 'intelligent behaviour' and partly the child's 'acculturation' into the 'anglo', upper-middle-class culture. She proposed that the test could still be used to estimate a minority child's learning capacity (intelligence), if learning opportunities could be held statistically constant. To achieve this, she developed different WISC-R scoring norms for different sociocultural groups. The groups were defined according to a multitude of criteria, including ethnic/racial background, family size, family structure, urban acculturation, and socioeconomic status. Different group norms were developed through multiple regression. By comparing the individual child's score, not with the national average, but only with the average performance of people from the same socio-cultural background,

---

7 Robert L. Williams, *The BITCH-100: A Culture-Specific Test*. Paper presented at the American Psychological Association Annual Convention (Honolulu, Hawaii, September 1972).

8 Wesman, 'Intelligent testing', p. 270.

Mercer argued that it was possible to estimate the learning potential of the child.<sup>9</sup>

Mercer's system was based on an explicitly formulated normative view of nationhood and culture. She favored cultural 'pluralism' and 'cultural democracy', and dismissed the 'conformity model', on which she argued both the education system and traditional testing were based. The content and language of the school curriculum and IQ tests were specific to the 'anglo core culture'. Minority children were thus assessed against the dominant culture. This led to a disproportionate number of minority children being assigned to classes for the intellectually disabled, and very few to classes for the gifted. Mercer held instead that the goal of public education should be to develop bilingual, bicultural, and 'bicognitive' children who could operate effectively both in minority and "core" culture.<sup>10</sup>

A number of arguments by Kaufman, Williams, Wesman and Mercer are relevant for the purpose of this chapter. Firstly, the argument that the WISC (and IQ tests in general) is a test of learned behavior, and therefore is a culture-specific test, adapted to a specific social group. Secondly, while Wesman dismissed the possibility of measuring universal constructs 'inside the child's head', Williams and Mercer maintained that this was possible, either through culture-specific tests adapted to the child's own culture (Williams), or through norms adjusted to the specific socio-cultural category to which the child belongs (Mercer). Thirdly, the idea that the relevant socio-cultural differences that must be "neutralized" in order to identify the child's true potential, are not only distributed along ethnic or "racial" lines, but also according to a number of socio-economic parameters. Fourthly, how one delineates, defines, and samples the population to which the test is adapted is as much a normative, political, and practical question as it is a methodological one. Cultural norms are built into the Wechsler tests; they measure how well a child is adapted to the society's "core" culture, which also permeates the school system. Let us consider these points as we now turn to the Norwegian WISC-R.

---

9 Jane R. Mercer, *Cultural Diversity, Mental Retardation, and Assessment: The Case for Nonlabeling*. Paper presented at the Fourth International Congress of the International Association for the Scientific Study of Mental Deficiency. (Washington, D.C., August 22–27, 1976).

10 Mercer, *Cultural Diversity*, p. 6.

## The Background of the Norwegian WISC-R Adaptation

In 1974, the Norwegian Fund for General Research (NAVF) established an expert committee to assess the need for a 'pedagogical-psychological instrument service', and for NAVF funding of research in this field. The initiator and leader of this committee (hereafter called the EPS Instrument Committee) was Hans Jørgen Gjessing, former director of *Statens spesiallærerhøgskole* [The State College of Special Pedagogy], established 1963, who became professor of educational psychology at the University of Bergen in 1972 and was a key figure in the professionalization of special education in Norway. The committee arranged a survey among employees in the Norwegian Education Psychology Services (EPS) and a conference with psychologists, medical doctors, pedagogues, and optometrists. It concluded that despite the twenty-five-year-long existence of the EPS, hardly any research had been done in the field. This also applied to the development of 'diagnostic aids', described as a decisive obstacle for the activities of the EPS.

A key task of the EPS was to give advice about special education for children with reduced learning abilities. From 1955, all Norwegian municipalities were mandated by law to provide special needs education, and since then, the state had co-funded municipal school psychologist posts that aimed to examine children with extra pedagogical needs (see Ludvigsen, chapter 6, this volume). The number of children receiving special education measures had risen from a few hundred in 1955 to about 45,000 in 1974. While acknowledging the necessity of many of these measures, many EPS employees saw this increase as problematic. Exaggerated use of resources on special education could preclude more appropriate pedagogical measures in the normal school, and special education could be pointless or even harmful for the individual student. The problem was a diminishing ability to identify the students who would really benefit from special education. In addition to this, the EPS had recently been given new statutory duties towards both pre-school children and high school students. There was therefore a need for more and better 'diagnostic aids in a number of areas'.<sup>11</sup>

11 *Diagnostiske hjelpemidler i den pedagogisk-psykologiske rådgivningstjenesten. Innstilling avlevert November 1974*, pp. 4–5. The committee also pointed out that the separation of weak pupils from the normal classes could lead to reduced tolerance for the "deviant" child.

The diagnostic activities of the EPS—testing in particular—had faced criticism for a long time. As demonstrated by Haave and Kyllingstad, and Thue (chapters 4 and 11, this volume), already the 1950s saw the rise of a wave of critique directed at testing and test-based pedagogical research and practice. The EPS Instrument Committee acknowledged that some of the critique was justified. Testing had, historically, been used uncritically as a ‘selection instrument’ for admission to special schools, with little focus on ‘the remedial effect’. But the committee claimed that tests still had a role to play; to enable early intervention in learning disabilities, it was important to develop ‘diagnostic tests with a high remedial value’. There was a need for new instruments to diagnose hearing, seeing, reading, writing, and motor impairments, as well as tests for cognitive functions. General ability tests were assigned a key role. A general ability test combined with a specific ability test could clarify whether a child had a general intellectual disability or a specific functional impairment. A properly normed IQ test could be a reference point for the interpretation of other tests that were not normed for Norway.<sup>12</sup> With limited resources, the committee proposed prioritizing the adaptation of international tests that were psychometrically validated, and to which significant qualitative information was attached. They suggested a more or less complete standardization of the WISC-R.<sup>13</sup> The WISC test’s huge international impact and distribution was thus an important argument for developing a Norwegian version.

The only general ability test that was available in an official Norwegian version at this point was the 1954 version of Terman–Merrill’s 1937 version of Stanford–Binet, which, for many decades, had been the internationally leading IQ test (see Paulsen, chapter 9, this volume). However, the EPS Instrument Committee considered the Norwegian version not only to be outdated, but also of generally low quality, with norms based exclusively on schoolchildren from Oslo. They also argued that the Stanford–Binet test produced little information beyond a total IQ score, and that the IQ score itself was of limited value for the assessment of remedial measures. The WISC test, in contrast, produced more information about the specific abilities of the child, was more interesting for the test taker, more flexible in

---

12 Diagnostiske hjelpemidler, p. 6.

13 Ibid., pp. 7–8.

use, and provided a better starting point for qualitative interpretation. It was, on the whole, more relevant for the assessment of remedial measures, and the fact that unauthorized and non-normed WISC translations were already in widespread use demonstrated that psychologists saw it as a better alternative.<sup>14</sup>

A key argument for prioritizing WISC was that it produced more information than just an IQ score. WISC-R is made up of subtests that measure different aspects of intelligence. In addition to the full-scale IQ, it produces a performance IQ and verbal IQ, as well as twelve primary scores. By comparing this data, an individual's cognitive strengths and weaknesses could be identified. Wechsler saw this as a useful procedure for the assessment and diagnosis of individuals, and Kaufman's book *Intelligent Testing* demonstrated how such interpretations of an individual's test results could lay the groundwork for the development of hypotheses about pedagogical needs and psychiatric conditions. It was due to this aspect of the WISC test that the committee saw it as more useful for the assessment of 'remedial measures' than the existing Stanford-Binet.<sup>15</sup>

The EPS Instrument Committee's report did not include a precise definition of 'intelligence' or 'general ability', but it seems to have been based on implicit ideas close to Wesman's (1967) viewpoint, namely that an IQ test probes a wide array of earlier learning experiences, which include not only formal education but also skills and knowledge achieved through life experiences more broadly. The committee made a distinction between intelligence tests and pedagogical tests. While pedagogical tests 'evaluate[d] or diagnose[d] the results of a more formal learning and education activity', intelligence tests measured cognitive abilities 'of a certain stability', which, to a lesser degree, were the result of formal learning. In line with this, the Committee argued, for example, that IQ testing could detect 'underachieving' students by disclosing 'a mismatch between general intelligence and educational achievements'. However, they also underscored that it was 'no longer' possible to draw a sharp distinction between abilities and learned skills.<sup>16</sup>

---

14 Ibid., p. 16.

15 Ibid., p. 16.

16 Ibid., p. 14.

This line of argument seems to suggest that intelligence tests are culturally loaded, but the issue was not thematized. Ongoing American controversies about the in-built cultural bias of IQ tests were not made relevant to the question of transforming an American test into a Norwegian one. Test adaptation was generally handled as a purely technical issue without taking normative aspects into consideration.

When the EPS Instrument Committee launched its report in late 1974, NAVF had already received and accepted a proposal for the funding of a Norwegian WISC-R standardization.<sup>17</sup> The project was anchored at *Norges Lærerhøgskole* [The Norwegian Teachers College], which recently (1968) had merged with the University of Trondheim and established a Psychology Department that offered undergraduate courses in psychology, differential psychology and psychometrics. Intelligence research was among the main pillars of the undergraduate courses and the professional community that formed at the new department. Applications for NAVF-funded research projects were part of a strategy for recruiting young scholars to the new department.<sup>18</sup>

Two of the expert members of the EPS Instrument Committee, Bjørn Foros and Johan Olav Undheim, had recently handed in their *magistergradsavhandling* [MPhil dissertations] at the new Psychology Department. To explore current theories about the “structure” of intelligence, they had undertaken factor analytic studies of test result data from Stanford–Binet testing (Foros) and WISC testing (Undheim) of schoolchildren. The WISC-R adaptation project was initiated and led by Undheim. As part of his research, he began to adapt a Norwegian version of the original 1949 WISC edition,<sup>19</sup> but when he became aware of the imminent launching of the revised edition, he decided to start

17 The archive of The Norwegian Fund for General Research is owned by Riksarkivet [The National Archives of Norway] and have the archival reference RA/S-2918, NAVF, Rådet for samfunnsvitenskapelig forskning, hereafter referred to as NAVF-archive. NAVF-archive, box B.65.01-015, 12.08.1974, Til Undheim, Rådet godkjenner prosjektendringen og de nye planene and box B.65.01-015, 19.08.74 (j nr. 7302) UNIT Psyk inst v/Undheim, Norsk tilpasning av WISC-R testen.

18 NAVF-archive, box Dche 65, Universitetet i Trondheim NLH 1968–1988, box Dche 0426, 10.07.1968, Hilmar Nordvik, NLH: Psykologisk institutt, forespørsel om betingelser for studentstipend i forb. med utbyggingen av forskningsmiljø ved Psyk. inst.

19 This was funded by NAVF and Åndssvakesakens Forskningsfond, and undertaken under the supervision of Undheim’s MPhil supervisor Hilmar Nordvik.

from scratch with WISC-R.<sup>20</sup>

In his proposal for NAVF funding, he pointed to the ongoing work of the EPS Instrument Committee and put forward arguments similar to those presented in the committee's report. Undheim maintained, for example, that most special education measures presupposed knowledge about the child's 'cognitive resources'; that there existed no adequate Norwegian test for that purpose; that a Wechsler test provided both a general intelligence score and an ability profile; and that it was the most widely used test internationally.<sup>21</sup>

## The Implementation of the Norwegian WISC-R Project



Fig. 10.1 WISC-R test kit. Photo: Helge Brekke/Museum of University History, KHM. University of Oslo. All rights reserved.

- 20 NAVF-archive: box Dche 65 Universitetet i Trondheim NLHY 1968–1988, box Dche 0426 Hilmar Nordvik 13.05.1971, box B.65. 01-3 WISC: søknadsskjema and 11.05.1971 WISC: søknad, vedlegg, (prosjektskisse) from Nordvik (NLH: psyk. inst) to NAVF, box B.65.01-015, 28.01.1974, Johan Olav Undheim p.v.a. Psyk. Inst. UNIT, NLHT, søknad for 1976, J.nr 1135, box B. 65.01-15 UNI Psyk. Inst. NLHT v/ Undheim til NAVF, Ang WISC-prosjektet, vesentlige forutsetninger for prosjektet er endret, box B.65.01-015, 21.05.1975, Johan Olav Undheim p.v.a. Psyk. Inst. UNIT, NLHT, Søknad for 1976, Søknadsskjema for 1976.
- 21 The Wechsler test was, furthermore, described as necessary basic equipment within educational counseling and clinical psychology, developmental psychology, brain damage diagnosis, and intellectual disability research: NAVF archive, box B.65.01-015, Johan Olav Undheim p.v.a. Psyk.inst. UNIT, NLHT, 15.05.1973, Søknad for 1976, box B.65.01-15 Søknadsskjema 'Norsk tilpasning av WISC-testen', søknad for 1974, 15.05.1973, box B.65.01-15, Ad søknad om midler til en norsk tilpasning av Wisc. XXX The proposal was supported by the Norwegian Psychologist Association: NAVF archive, box B.65.01-015, 28.06.1973, jnr 4660 Norsk psykologforening til NAVF vedr. standardisering av WISC-testen for norske forhold.





Fig. 10.2 WISC-R block design. Photo by Jon R. Kyllingstad. CC BY-NC.



Fig. 10.3 WISC-R picture completion. Photo by Jon R. Kyllingstad. CC BY-NC.



Fig. 10.4 WISC (1949) Picture arrangement. Photo: Helge Brekke/Museum of University History, KHM, University of Oslo. All rights reserved.

WISC-R consists of a set of so-called 'performance' and 'verbal' tests. The 'performance' tests included 'picture completion' (to add a missing element to a drawing), 'picture arrangement' (to arrange a series of illustrations into a meaningful comic strip), 'object assembly' (a jigsaw puzzle), 'block design', 'coding', and 'mazes'. The verbal tasks included 'arithmetic', 'digit span', 'information' (a test of factual knowledge), 'similarities' (with questions such as, 'In what way are a piano and a guitar similar?'), 'vocabulary' (with questions such as, 'What does "grumpy" mean?') and 'comprehension' (with questions such as, 'Why do we have to pay taxes?'). The verbal test scores and the performance test scores form the basis for the calculation of, respectively, a Verbal IQ, and a Performance IQ, and the two taken together give the Full Scale IQ.<sup>22</sup>

The scores indicate how well an individual schoolchild performs compared to the average of their age group. The procedure for calculating IQ scores is based on the premise that IQ is normally distributed around

22 Johan Olav Undheim and David Wechsler, *Håndbok WISC-R: Wechsler Intelligence Scale for Children—Revised*, Norwegian edition by Johan Olav Undheim (Oslo: Norsk psykologforening, 1978).

a mean of one hundred, and with a standard deviation of fifteen. Test results are therefore, by definition, distributed in line with a bell-shaped, normal distribution curve. The IQ scale is nothing but a tool to compare the intelligence of a single individual with a total population, and as David Wechsler himself emphasized, the IQ scale is only valid for the specific group for which the scale is constructed.

To transform the American WISC-R into a Norwegian one, test instructions and verbal tests were translated. The tasks and the order of tasks were then adapted to the new cultural frame of reference, and the test was 'normed'. This refers to the mathematical procedure for converting raw test scores into composite index scores (Verbal IQ and Performance IQ) and a Full Scale IQ score, (re-)calibrated in such a way that the test results in the total Norwegian schoolchild population were normally distributed around a mean of a hundred.

The undertaking of these tasks began in 1974 with translation and cultural adaptation through 'qualitative trials' on 50 children, followed up in 1975 with 'systematic trials' of a preliminary test version on 200 children. The cultural adaptation aimed to ensure that the tasks had the same level of difficulty for Norwegian schoolchildren as the American original had for American schoolchildren, and that they were arranged in the same order of increasing difficulty. This led to some small, but significant changes in the performance tests and a great number of changes in the verbal subtests—'information', 'similarities', 'reasoning', and 'vocabulary'.<sup>23</sup>

Based on the 'qualitative' and 'systematic trials', a preliminary version was developed and then administered to 766 schoolchildren to gather scoring data for the norming of the scale and to fine tune the content of the subtests. This part of the project included collaboration with psychologists at about fifty EPS offices, who tested children at a number of schools. The schools were selected according to a plan for approximate stratified random sampling based on assumptions about the composition of the school's student body.<sup>24</sup> The criteria of

---

23 NAVF archive, box B.65.01-15, 'Norsk tilpasning av den reviderte WISC testen Arbeidsplan. To find adequate Norwegian terms, they tried, according to Undheim, to find words that were common among ordinary people, and which the child would encounter at school, at home, and in mass media.

24 NAVF archive, box B.65.01-015, 14.04.1977 (NAVF 15.04.77) Beretning fra Undheim til NAVF.

representativeness were language, county, urban/rural residence, and parents' occupation. The testees were drawn randomly from the chosen schools among children evenly distributed between the relevant age cohorts (6.5 to 15.5 years) and between boys and girls.<sup>25</sup>

The sampling criteria partly resembled the sampling criteria in the American WISC-R standardization, with the important exceptions of race and language. In contrast to the first American WISC version (1949), which had been normed on a purely white population sample, WISC-R was normed on a sample with representative proportions of two racial groups, so-called whites and non-whites, the latter group including 'Blacks', 'American Indians', 'Orientals', 'Puerto Ricans', and 'Chicanos', insofar as they appeared "non-white" based on 'visible physical characteristics'. This means that all non-white individuals were lumped together into one racial group based on a biological understanding of 'race'.<sup>26</sup>

'Race' was not a criterion in the Norwegian standardization. This is not surprising, given that the concept of race had a very different status in Norway than in the USA. In contrast to the early twentieth century, when Norwegian academics had been involved in the racialization of minorities and local populations in Norway, it was no longer common to classify Norwegian citizens into races in the early 1970s. Though race was still seen as a biologically meaningful concept even in Norway, neither a biological nor a socio-cultural concept of race was used in Norwegian statistics, social scientific research, or in the public discourse about Norwegian society.<sup>27</sup>

The US WISC-R sampling strategy included language as a criterion in the sense that only English-speaking children were included. Sociocultural language or dialect differences within the English-speaking population, such as the ones thematized by Williams and

25 Håndbok WISC-R. WISC-R was meant for 6–16-year-olds, but the sample only included children at the age of 6.5, 7.5, 8.5, 9.5, 10.5, 11.5, 13.5, and 15.5 years. The rationale was to save work and money and that it was most important with a reliable scale for the small children at the start of their school career and for children at the transition from primary to secondary school. Norms for the missing age groups were estimated based on the data from these age groups.

26 David Wechsler, *WISC-R: Wechsler Intelligence Scale for Children—Revised* (New York: The Psychological Corporation, 1974), p. 17.

27 Jon Røyne Kyllingstad, 'The absence of race in Norway?', *Journal of Anthropological Sciences*, JASs forum, Vol. 95 (2017), 1–9.

the BITCH-100 test, were not taken into consideration. The Norwegian sample, in contrast, was constructed to include representative shares of children that were taught either of the two officially acknowledged forms of Norwegian—*bokmål* and *nynorsk*—at school. The existence of two standard Norwegian languages has its historical roots in the four centuries of Danish rule over Norway that ended in 1814. The nineteenth century saw the rise of two competing principles of language standardization and cultural nation-building; one based on existing Norwegian dialects, the other on the Norwegianization of the existing Danish written language. The resulting two forms of Norwegian have different orthography and some vocabulary differences, but are mutually intelligible to a very high degree. The point was to find out whether children from *nynorsk* classes suffered a disadvantage, since the test was translated into *bokmål*. The conclusion was that no such disadvantage could be detected in the test results. The test manual underscores, however, that dialect differences might affect the reliability of the test, and that verbal instructions should be given in a language close to the child's dialect.<sup>28</sup>

Beyond this, linguistic or ethnic diversity was not taken into consideration in the linguistic and cultural adaptation and norming of the test, or in the writing of the manual. Neither the Sámi—an indigenous people of northern Scandinavia, speaking finno-ugric languages—nor the Kven, an originally finno-ugric speaking national minority, or any other minority groups were discussed in relation to representativeness and the area of validity of the test. Contrary to the stated goal of geographical representation, no children from Finnmark, the country's northernmost and most multicultural county, with the highest proportion of Sámi and Kvens, were involved in the project. Oslo, with the highest proportion of immigrants, was underrepresented. The Norwegian test manual does not even mention ethnic minorities, and to what extent and in what way minority status had to be taken into consideration in the testing of a schoolchild.

In contrast to the US sampling strategy, where the urban/rural distinction was operationalized as communities with more or less than 2,500 inhabitants, this distinction was, in the Norwegian version, defined as cities and industrial societies versus fishing or agricultural

---

28 Håndbok WISC-R, pp. VI, 2,19.

societies. Why this specific definition of urban/rural was chosen is not obvious, but one likely reason is to be found in existing research on rural/urban IQ differences. According to a contemporary psychology textbook, it was generally held that the average IQ was lower in rural communities compared to cities.<sup>29</sup> In 1961, this view had been confirmed in a large Norwegian survey that aimed to shed light on how learning outcome and intelligence were affected by the different organization of teaching within large urban schools and small rural schools with age-mixed classes. The conclusion was that schoolchildren and military conscripts from the countryside did not perform as well as children and conscripts from cities on tasks that demanded abstract, theoretical thinking and learning.<sup>30</sup> In 1961, a research paper was published to this effect, showing that Norwegian military conscripts with backgrounds in fishing and/or agricultural communities had, on average, a lower education level and lower IQ scores compared to both industry-based rural communities and cities.<sup>31</sup> The test scores collected by the WISC-R project in the mid-1970s, in contrast, showed no significant difference along the rural/urban divide.<sup>32</sup>

However, neither geographic nor urban/rural representation received much attention in the preparations for the project, or in the test manual's description of the project. This is in contrast to the substantial amount of attention given to the issue of socio-economic status. Pointing to his own MPhil dissertation, Undheim assumed that WISC scores and 'the ability level of the child usually co-varies with a number of socio-economic variables'. A socio-economically biased sample would thus affect the norming of the IQ scale and had to be controlled. Based on the occupation of the family's breadwinner, the children were given what Undheim described as 'a relatively traditional socio-economic ranking

29 Eva Nordland, *Innføring i psykologien* (Oslo: Aschehoug 1955), p. 132.

30 Vidkun Coucheron Thrane: *Avgangskarakterer i folkeskolen: Militærpsykologiske meddelelser no. k-1* (Oslo: Forsvarets psykologiske avdeling 1961), p. 21, etc. The study was meant to assess how the differently organized teaching in small rural schools and large urban schools affected the learning outcome and whether the grading system was fair. The conclusion was that the average child on the countryside were favored, receiving, on average, better grades while performing less well on the psychometric test than the city children.

31 Natalie Ramsøy, 'Evner, utdanning og yrkesvalg i norsk samfunnsstruktur', *Tidsskrift for samfunnsforskning* (1961), 217–237.

32 Håndbok WISC-R.

from one to four'. This socioeconomic ranking was found to correlate 0.3 with the total score on the WISC-R among the children who participated in the study.<sup>33</sup>

## Undheim's View on Intelligence and Culture

Alongside the WISC-R standardization project, Undheim also embarked on a doctoral project. He performed factorial analysis of data from the WISC and WISC-R testing of Norwegian children. In common with his MPhil work, the doctoral project aimed to assess current theories about the "structure" of intelligence. He focused mainly on Joy Paul Guilford's 'Structure of Intellect' theory, which rejected the notion of intelligence as a singular construct and claimed that intelligence was made up of a large number of abilities organized in three dimensions: operations, content, and products. He also assessed the theory of fluid and crystallized intelligence, put forward by the American psychologist Raymond B. Cattell and developed further by psychologist John Horn, whom Undheim befriended. Both his doctoral thesis and the WISC-R standardization work were partly undertaken during a research stay with Horn at the University of Denver.<sup>34</sup>

Undheim ended up dismissing the idea 'that statistically identified factors were "basic" entities, as argued by Guilford and others'. He argued instead for the restoration of general intelligence, Spearman's 'g', claiming that all entities identified through factor analysis were to a large degree correlated to 'g', and that 'g' had demonstrated an 'impressive predictive power [...] for almost all kinds of real life learning'.<sup>35</sup>

But what was 'general intelligence', according to Undheim? He argued that intelligence as measured by IQ tests was not a 'capacity' or 'potential', but a measure of 'different achievements'. As he wrote, 'the discrepancy between the general intelligence factor and school achievement' was not one between native potential and achievement, but rather 'a discrepancy between a very broad achievement index and a more narrowly defined academic achievement index'. This meant,

33 Håndbok WISC-R, p. 2–3. NAVF archive, box B.65.01-15 Johan Olav Undheim p.v.a. psyk inst unit NLHV søknad for 1977 30.09.1976 Undheim to SSB.

34 NAVF archive, box B. 65.01-11 Undheim Johan, Søknad for 1974 Reisebrev.

35 Johan Olav Undheim, 'On Intelligence IV: Toward a restoration of general intelligence', *Scandinavian Journal of Psychology* (1981), 251–265.

for example, that a person with a high IQ score and low scores on school achievement measures was not necessarily an underachiever with unused inborn potential. It could rather mean that the student's general intelligence performance was developed through 'wide reading and other intellectual pursuits, while specific school-related material is shunned'.<sup>36</sup>

In a 1988 paper, Undheim and his colleague Knut A. Hagtvet discussed the cultural specificity of IQ tests, based on Norwegian experiences with the translation and adaptation of intelligence tests, WISC-R in particular. They argued that such adaptations were generally based on the implicit assumption that IQ measures are universally valid within western industrialized countries due to 'the presence of certain common cultural factors'. The paper attempted to examine this assumption and was partly an elaboration of arguments that Undheim had made ten years earlier in the WISC-R test manual.<sup>37</sup>

Undheim and Hagtvet found that IQ scores in Norway were correlated with school grades, teacher ratings, achievement test scores, and indices of socioeconomic level and upward social mobility, and that this paralleled similar findings in the USA and other western countries. They also argued that the factor structure obtained through the analysis of Norwegian WISC-R test results resembled the factor structure obtained in the original American standardization samples. This suggested, according to Undheim and Hagtvet, that the test measured the same thing in Norway as it did in the USA (and in other western countries), and that "intelligence" was structured the same way across western populations. This did, however, not necessarily mean that it was possible to do a simple 'metric comparison of the level of intellectual performance between (western) nations'—that is, to compare test results from, say, Norway and USA in order to find out which had the most intelligent population.<sup>38</sup>

It was obviously problematic to compare the verbal tests. But even the performance test results were not directly comparable, according to Undheim

36 Undheim, 'On Intelligence IV', p. 261.

37 Knut A. Hagtvet and Johan O. Undheim, 'The Norwegian Experience of Test Use: A Selective Review of Norwegian Tests and Measurements in Cultural Context', in *Human Abilities in Cultural Context*, 1st ed., ed. by S. H. Irvine and J. W. Berry (Cambridge: Cambridge University Press 1988), pp. 263–81.

38 Undheim and Hagtvet, 'The Norwegian experience', p. 271.



and Hagtvet. It had been necessary to change some of the performance tests in the Norwegian WISC-R adaptation. A playing card image was, for example, deleted from the Picture Completion test on the grounds that playing cards were less widespread in Norway and could evoke negative feelings in some religious groups. Similarly, a picture series of a boxing match was deleted from the Picture Completion test, since Norwegian children had little knowledge about boxing.<sup>39</sup>

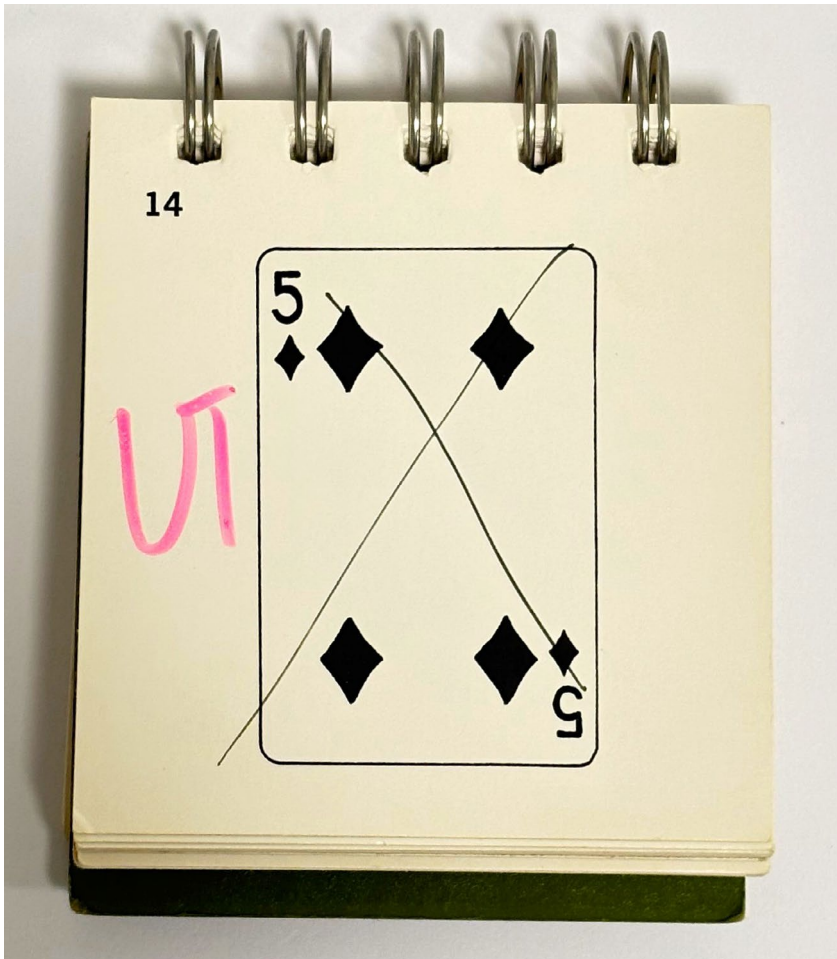


Fig. 10.5 Even if this playing card was not included in the Norwegian WISC-R version, the test kit was imported unmodified. Photo: Jon R. Kyllingstad. CC BY-NC.

<sup>39</sup> Undheim and Hagtvet, 'The Norwegian experience', pp. 272–273; *Håndbok WISC-R*, p. 5.

Only five of the twelve WISC-R subtests had been kept identical across the US and Norwegian versions—namely block design, object assembly, coding, digit span, and mazes—but Undheim and Hagtvet still argued that it would require an 'analysis of child culture' to determine whether these tests were 'culture fair' to Norwegian children. Children's experience with test-similar material could even change over time. Play kits similar to Wechsler's block design test had, for example, only recently been observed in Norwegian toy stores.<sup>40</sup>

In addition to the issue of cross-national comparability of the very test items, they also problematized the comparability of the population sample on which the test was normed. As we have seen, there were some significant differences between the sampling procedures in the US and Norwegian WISC-R standardizations, and Undheim and Hagtvet acknowledged that such procedures would always differ, since institutional and other circumstances differed between countries. Being aware of these problems, Undheim and Hagtvet still found it fruitful to compare US and Norwegian performance test scores. American children scored on average significantly better than Norwegians on digit span and coding, while Norwegian children had considerably higher test scores on the other five performance tests. Four of these were particularly strongly correlated to Full Scale IQ, and this indicated, according to Undheim and Hagtvet, a significantly higher average 'general intelligence' in Norway than in the USA.<sup>41</sup>

They proposed several possible explanations. One obvious explanation was that 'disadvantaged groups of non-whites' contributed to lower the average American Performance IQ. But they also hypothesized that the Norwegian school system was, generally, 'more efficient in delivering the skills that lead to effective test performance', and that certain aspects of Norwegian society were positive for school efficiency. This included more teachers per student, homogeneity of religion and language, few foreign residents, and a low immigration rate, as well as low rates of aggressive violence, industrial disputes, unemployment, and divorce, high numbers of published book titles and newspapers, and volumes in public libraries per capita. All these aspects put Norwegian children 'in an extremely advantageous position',

40 Undheim and Hagtvet, 'The Norwegian experience', p. 272.

41 Ibid., p. 273.

compared to American children.<sup>42</sup>

Finally, both the 1988 paper and the WISC-R test manual pointed out that there was a greater degree of variance in the American test results than in the Norwegian. The proposed explanation was that Norway had a very high level of general education, was more culturally and linguistically homogeneous, and had a more egalitarian distribution of wealth, health and education than the USA. In the test manual Undheim claimed, however, that given the role of 'socio-cultural and nutritional conditions' for 'the development of the abilities of children', he would have expected an even greater difference between the two countries. He had expected that the socially, linguistically, and culturally homogeneous Norwegian population would produce more homogeneous test scores than it actually did.<sup>43</sup>

### Culture, Intelligence, Class and Norwegianness in the Norwegian WISC-R

The discussions leading up to the WISC-R standardization project, the standardization project itself, and Undheim's analysis of test score data from the project, indicate that the undertaking was based on the insight that WISC-R is a culturally loaded test. Intelligence, as measured with an IQ test, was identified as a "Western" construct, consisting of measurable components that are unevenly distributed between Western nations and between socially and culturally defined groups within nations due to social, economic, and cultural differences. If we are to take this insight at face value, it means that there was nothing in the project's theoretical and methodological framework or empirical findings that itself justified the lumping of Norwegian school children into one

---

42 Undheim and Hagtvet, 'The Norwegian experience', p 274. They took into consideration the minor changes that had been made to two of the WISC-R performance tests, and they proposed sociocultural explanations of the test score differences where the American children performed best, coding and digit span. The coding test 'depends on a careful trade-off between speed and accuracy' in carrying out an 'intellectually easy, repetitive task', and US children may have been 'more practiced in finding the optimal balance between speed and accuracy'. This was due to the widespread use of 'multiple choice formats and the like in school', and that 'the United States intellectual "climate" is generally more rewarding of the quick-and-fast answer than most European countries'.

43 Håndbok WISC-R, p. 13–14.

group and the development of one joint IQ scale for all members of that group. It would, for example, have been equally reasonable to seek inspiration from Mercer's SOMPA-project, and develop different norms for different socially defined groups of children. This option was never on the table, as the basic aim of the project was to construct one test and one national norm which all school children could be measured against. This aim was mainly dictated by political and institutional conditions; the national school system's need for a tool to solve a practical problem, and the available resources in terms of research funding, as well as the institutional infrastructure and personnel in the EPS.

It is also interesting to note that the question of the representativeness of the population sample was mainly discussed in relation to the norming of the scale, and less in relation to the cultural adaptation of the content of the test. If WISC-R was a culturally loaded test, the important question is, of course, how one defines and delineates the "culture" against which the individual child is measured. This is, in turn, a question of how to identify the group of people that are chosen to represent the culture to which the test is adapted. A possible strategy could have been to seek inspiration in Williams's BITCH-100 test and develop different tests for children with different socio-cultural backgrounds. This was never an issue. The test was culturally adapted to Norwegian culture which, as a result, was treated as one, unified entity. The project was thus based on the implicit premise that all Norwegian schoolchildren share the same culture, and that this culture is a uniform entity.

It can also be argued that if the ideal had been to create a "culture-fair" test that was equally fair to Norwegian children from different socio-cultural and ethnic-cultural backgrounds, it would have been important to strive to include only test tasks that were equally difficult for children from these different backgrounds. This was, however, not on the agenda in the WISC-R project. Regarding socio-economic differences, the project ended up creating a test where the average child from a low socio-economic background scored less than the average child from a high socio-economic background. This could be taken to mean that the test was biased against the cultural frame of reference of children from low socio-economic backgrounds, but Undheim did not see this as a lack of cultural fairness. Instead, he took it to confirm the validity of the test, since the statistical correlation between social class

and test scores in Norway coincided with a pattern that was common to Western nations in general. The standardization process was based on, and confirmed in a circular move, the presupposition that measured IQ is unevenly distributed among social classes.

As the IQ score was considered a measure of skills and knowledge achieved through both formal education and life experiences more broadly, this meant that the test was actually constructed as an instrument to measure the child's ability to function within a socio-cultural context defined by the children of parents with a high socio-economic status.

While much attention was given to socio-economic differences in the sampling strategy and the discussion of the test score data, the lack of attention to linguistic and ethnic differences is striking. It is reasonable to argue, as Undheim did, that the American population was much more heterogeneous than the Norwegian was. The two countries were very different with regard to the size of territory and population, as well as degree of political-administrative, ethnic, social, and linguistic heterogeneity. It is still striking that the process of transforming an American test into a Norwegian one was more or less unaffected by concepts, topics, and insights from the contemporary American debates about racial, ethnic, linguistic, and socio-cultural bias and fairness in IQ testing.

The methodological issues in the standardization project were generally discussed in a "technical" way, not taking into consideration normative questions about nationhood, cultural pluralism, and democracy. It is still reasonable to argue that the project was implicitly based on what Mercer defined as a 'conformity model', and that the lack of attention to questions about cultural diversity within the Norwegian population may have been due to a notion of Norway as a culturally homogeneous society. This notion coincided with dominating attitudes in the Norwegian public. It was also in line with the approach to cultural and linguistic differences that characterized the national curriculum guidelines for the elementary and secondary school, *Mønsterplanen for grunnskolen*, adopted by the Norwegian parliament in 1974. According to this, non-Norwegian-speaking children were entitled to special training in Norwegian as a foreign language. But they had no right to

education in their own mother tongue, and concepts such as Mercer's bilingual or bicultural children had no place in the plan. The only, but important, exception to this rule, was that Sámi children from so-called language-mixed communities had the right to education in their native language.<sup>44</sup>

The curriculum guideline's provision on Sámi languages was meant to counteract the effects of a century-long harsh assimilation policy against the Sámi by the Norwegian state. From the mid-nineteenth to the mid-twentieth century, the Norwegian school system had been mandated to work for the "Norwegianization" of the Sámi, with the aim of replacing Sámi language and identity with Norwegian language and identity. This policy was officially abandoned in 1963, though it took many years before the actual teaching practices changed. According to the 1974 guidelines, however, the schools were explicitly mandated to help Sámi children become bilingual and bicultural in order to maintain Sámi language and culture. The curriculum guidelines also discussed the pedagogical challenges involved in the meeting between non-Norwegian-speaking Sámi children and a mainly Norwegian school environment.<sup>45</sup> With this background, it is striking that the Sámi were not even mentioned in connection with the WISC-R adaptation project. It is in particular significant that the question of the test's validity vis-à-vis Sámi (and other minority) children was omitted in the manual.

Children with immigrant backgrounds were also conspicuously absent from the discussions around the Norwegian WISC-R adaptation. Norway, with about four million inhabitants, received approximately 150–200,000 immigrants between 1966 and 1977.<sup>46</sup> According to the national curriculum guidelines of 1974, non-Norwegian-speaking immigrant children and Kven children had the right to language training for foreign speakers.<sup>47</sup> Beyond this, the guidelines do not mention these children. The issue was also completely absent from the WISC-R standardization project, and from the test manual.

44 *Mønsterplan for grunnskolen* (Oslo: Kirke og undervisningsdepartementet/Aschehoug & Co., 1974), pp. 71–72.

45 *Mønsterplan*, pp. 72, 125–126.

46 Statistical yearbook of Norway [Statistisk årbok] for the years 1968 to 1978: 'Immigrations' [Innflytting til landet]. The total immigration to Norway was 195,909.

47 *Mønsterplan*, p. 72,

## WISC-R and the Testing of Language-Minority Children

We will now leave aside the coming-into-being of the Norwegian WISC-R, and instead turn our attention towards the use of the test. Which notions of culture and intelligence guided the practical use of the WISC-R, and how did these notions correspond to the conceptualization upon which the development of the test had been based? A comprehensive investigation of these questions is well beyond the scope of this chapter, but it is still possible to shed some light on the issue. We begin with a brief look at the *Handbook for the Interpretation of WISC-R*, published in 1990 (and in an expanded version in 1999) as a companion to the test manual and mainly meant for employees in the EPS service. It was written by psychologist Hans Georg Kayser, but consisted mainly of selected notes from Alan Kaufman's book *Intelligent Testing with WISC-R*.<sup>48</sup>

The introductory chapter of the Norwegian 1990 handbook cites Kaufman's statement that IQ is not a measure of genetic potential, and that the WISC-R test measures behaviors that the child has learned, but it doesn't include Kaufman's point that learning takes place in a culture and the test therefore is 'culture loaded'.<sup>49</sup> The Norwegian handbook includes no references to American discussions about class, race, ethnicity, and cultural bias, in spite of the fact that Kaufman himself positioned his book at the center of these debates—as a middle ground between those who wanted a moratorium or abolishment of IQ testing in American schools, and those, like Arthur Jensen, who 'staunchly defend the importance of the g factor for blacks and whites'. Kaufman argued for the potential benefits of intelligence testing, but warned against 'abuse' and the 'equation of test scores with genetic potential', urging educational psychologists to be aware of the inaccuracies of WISC-R and other IQ tests, including problems of racial, social, and cultural fairness.<sup>50</sup> To interpret WISC-R test scores from minority children, the examiner needed specific insight into linguistic, cultural, and social issues, and for the assessment of 'Blacks, Chicanos, and working class whites', Kaufman recommended the use of Jane Mercer's SOMPA test battery, including the socio-culturally specific norms for the interpretation of WISC-R raw scores.<sup>51</sup>

48 Hans Georg Kayser, *Håndbok for tolkning av WISC-R* (Brandbu: Skolepsykologi - materiellservice, 1990).

49 Kayser, *Håndbok* (1990), p. 12.

50 Kaufman, *Intelligent Testing*, p. 18 etc.

51 Ibid., p. 16.

In contrast to these statements in the introductory chapter, the remainder of Kaufman's book pays far less attention to the practical implications of these issues, and in Kayser's selective notes from Kaufman, the topic is even more absent. In a book of over a hundred pages, the theme is allotted only fourteen lines of text, relating to the question of how to interpret a combination of low verbal IQ and high performance IQ scores. One explanation for such a discrepancy may, according to Kayser, be related to a child's language minority background. In such cases, the verbal score will not reflect the child's intellectual potential. Both the verbal and the total IQ score will be meaningless and should never be used in the diagnostic procedure. This means that, in contrast to the book's introductory statement which defines the IQ score as a measure of something that is learned, the IQ score is here interpreted as a measure of inborn potential, and the WISC-R test dismissed as useless on multilingual children. To the extent that "culture" is mentioned elsewhere, it is only in passing, and always to point out that a child's lack of 'cultural possibilities at home' may affect a test score.<sup>52</sup>

WISC-R became a key tool in the diagnostic toolbox of many EPS offices in a period (circa 1978–2003) when Norway became increasingly heterogeneous. In 2007, 415,000 of 4.6 million inhabitants were either immigrants or children of immigrants,<sup>53</sup> and as early as 1997, 27.5% of the schoolchildren in Oslo belonged to a linguistic minority. A substantial number of these children were in contact with the EPS during their school career. Children with an immigration background were strongly overrepresented among students that received special education measures.<sup>54</sup> The 2000s saw the publication of three studies which suggested that an important reason for this was the testing of language-minority children with the WISC-R, without taking into consideration that their linguistic and cultural background did not fit with the background of the children to which the test was adapted and normed.

In 2003, the pedagogue Joron Pihl published a discourse analysis of 125 assessment reports of minority children, written at two EPS offices in Oslo during the 1990s. The children were referred to the EPS because of language difficulties, academic difficulties, behavioral difficulties, or

---

52 Kayser, *Håndbok* (1990), pp. 14, 26–39, 40.

53 Statistical yearbook of Norway 2008.

54 Nordahl & Overland, *Idealer og realiteter*.



socio-emotional difficulties. In most cases, the EPS service had tested the child with the Norwegian WISC-R, and IQ scores often had decisive importance for the selection of pupils to special education. Pihl argued that the test was used as if it were culturally neutral in most of these cases, and test results were interpreted with an insufficient consideration of the child's level of competence in the Norwegian language.<sup>55</sup>

In the 2005 book *Det sakkyndige blikket* [The Expert's Gaze], Pihl criticized the EPS services in Oslo and put their contemporary test practice into a context that included the history of scientific racism, eugenic sterilization, and oppression of minorities in Norway. The book was welcomed by some in the professional community for pointing at real problems and raising an important discussion. But it also faced criticism from EPS employees who did not accept Pihl's description of the situation, objected to the study's qualitative methodology, or argued that the situation had changed for the better since the early 1990s.<sup>56</sup>

In the late 2000s, Pihl undertook a new study based on 125 assessment reports from the same EPS offices during the period 2001–2005. The content of these were analyzed with both quantitative methods and quantitative discourse analysis. She found many of the same patterns as in her previous study. Both studies demonstrated that minority children were most often assessed as if they were majority children, and that lacking Norwegian language skills was interpreted as a lower intelligence level.<sup>57</sup>

55 Joron Pihl, *Etnisk mangfold i skolen. Det sakkyndige blikket* (Oslo: Universitetsforlaget, 2010), pp. 107–137.

56 Newspaper articles about Pihl's research and op-eds by Pihl: 28.09.2002, 29.09.2002, *Halden Arbeiderblad* 28.09.2002, *Telemarksavisa* 30.09.2002, *Dagbladet* 31.10.2002, 27.11.2002, *Aftenposten* 24.08.2010, *Aftenposten Aften* 20.04.2005, 21.04.2005, 22.04.2005, 03.05.2005, 06.05.2005, 19.12.2007, 26.08.2010, 14.09.2010, *Dag og Tid* 07.02.2004, 23.04.2005, *Uttrop* 08.-15.02.2007. Critical newspaper op-eds about Pihl's research: Anne Messel, 'Ordkrigen om IQ-testingen', *Dagbladet* 14.11.2002, Kathe Lundahl, 'Vrengebilde av Pedagogisk-psykologisk Tjeneste (PPT) Aftenposten Aften 07.07.2005, Walter Frøyen, 'Begrepet IQ brukes ikke', *Aftenposten Aften* 02.09.2010. Børre Hansen: 'Hvem står for ufagligheten, Pihl?', *Aftenposten Aften* 12.05.2005. Astri Heen Wold 'Psykologi og etnisk mangfold i skolen', *Aftenposten Aften* 06.05.2005, Børre Hansen: 'Joron Pihl gjentar sin historie', *Aftenposten Aften* 20.04.2005. Articles in academic journals: Espen Egeberg: 'Bokomtale. Etnisk mangfold. Det sakkyndige blikket', *Spesialpedagogikk* no. 6 (2005), p. 42–43, Rolf Fasting, 'Kommentar til boka 'Etnisk mangfold i skolen. Det sakkyndige blikket' av Joron Pihl', *Spesialpedagogikk* no. 8 (2010) p. 47, Karianne Westrheim, 'Replikk', *Spesialpedagogikk* no. 10 (2010) s. 32. Astri Heen Wold: 'Kan for lite norsk, stempler som dum', *Tidsskrift for norsk psykologforening* 01.12.2006.

57 Pihl, *Etnisk mangfold*.

2010 also saw the publication of an extensive doctoral thesis by Knut Erik Aagaard, a psychologist who had been working in the Oslo EPS from 1974 to 1996. His project aimed both to describe and assess the quality of the EPS assessments of minority children, and to intervene in the field to help raise the quality of these assessments. His study of the established practice was based on the scrutiny of forty-one EPS reports and interviews with forty employees at EPS offices from the nine counties of south-east Norway in the period 1997–2004, as well as sixty reports from three EPS offices from 2007–2010.<sup>58</sup>

Aagaard's findings showed the same pattern as Pihl's. In general, the EPS advisors tended to assess language-minority students as if they belonged to the language majority, without sufficiently considering the relevant differences in language and culture. They lacked the necessary research-based competence in bilingualism and its pedagogical consequences, and did not, to a sufficient extent, investigate the children's language-learning history or conditions. Language-minority students were often tested in Norwegian with a WISC-R test (or its successor WISC-III), without taking into due consideration that the child's mother tongue was not Norwegian. According to Aagaard, some of the interviewees seemed totally unprepared for the question about Norwegian language proficiency, and showed a lack of knowledge or respect for the test's psychometric prerequisites.<sup>59</sup>

Both Aagaard and Pihl argued that unqualified use of the WISC-R test was part of a general pattern—the EPS service had weak competence to distinguish between learning difficulties and the consequences of unfavorable language-learning conditions. The language-minority child's pedagogical needs were too often assessed without taking into account the mismatch between the child's linguistic and cultural background and the monolingual school situation. The schoolchild's problem with the school situation was most often described as a problem within the child themselves, not as a problem that is caused by the school or other aspects of the child's surroundings.

Children with weak Norwegian language skills were legally entitled to three pedagogical measures: language education in their mother

---

58 Knut Erik Aagaard, *Den språklige faktor: Pedagogisk-Psykologisk utredning av barn med minoritetsspråklig bakgrunn* (PhD dissertation, Oslo: University of Oslo, 2010), pp. 421, 431, 487.

59 Aagaard, *Den språklige faktor*, p. 30, 357, 361, 373, 493.

tongue, bilingual teaching in regular school subjects, and education in Norwegian for non-native speakers. Instead of these language-focused measures, minority children were often categorized as having general learning difficulties or language difficulties and given special education measures intended for children with such difficulties. In this way, they were 'cheated of the effective pedagogy they could benefit from', according to Aagaard.<sup>60</sup>

Both studies point to how the misplacement of language minority children in special education exemplifies a general aspect of ethnocentrism in the Norwegian school system, characterized by an assimilatory approach lacking updated insight into research on the effects of bilingualism on a child's psychological development, and a devaluation of minority languages. In other words, the usage of WISC-R to select minority children for special education was part of a school system characterized by what Mercer defined as a 'conformity model' of nationhood.<sup>61</sup>

A key element in both Aagaard's and Pihl's critique was that the use of WISC-R on minority children was highly problematic, since these children did not belong to the group for which the test was adapted and normed. In violation of Kaufman's (and Kayser's) principles of intelligent testing, in practice, the IQ scores were often interpreted as a valid measure of the child's inherent intelligence, and used as an explanation of the child's weak ability to benefit from school education. But Aagaard and Pihl had different views on the usefulness of IQ testing. Pihl was generally critical of IQ testing, arguing that testing resulted in the problem being placed on the child's head instead of in the educational environment. She was also generally dismissive towards all use of ability tests on children that did not belong to the group the test was standardized for.<sup>62</sup> Aagaard, on the other hand, held that a Wechsler test could give useful information about language-minority children as well, if used in a correct way—that is, not seen as a measure of a child's given abilities, but as a starting point for developing hypotheses about causal factors.

For instance, the combination of a low Verbal IQ score and high Performance IQ score, it was argued, should be regarded as a normal

60 Aagaard, *Den språklige faktor*, p. 379, 386, 392, 493 (quote), etc.

61 Pihl, *Etnisk mangfold*, pp. 199–223, etc., Aagaard, *Den språklige faktor*, pp. 189, 209–210, etc.

62 Pihl, *Etnisk mangfold*, pp. 107–136.

cognitive profile for a child who does not fully master the language of the IQ test. Both the low verbal IQ score and a low full-scale IQ score would in such cases normally be the result of weak Norwegian language skills. However, the performance test could, in most cases, be used as an indication of non-verbal cognitive abilities in a language-minority child. Even the verbal part of the Wechsler tests was informative, according to Aagaard, if interpreted correctly. If, for example, a minority-language speaker scored better on verbal abstraction (the similarities test) than on the vocabulary test, this could indicate that the problem lay in poor Norwegian vocabulary rather than in ability level.<sup>63</sup>

Aagaard also countered a claim from Pihl that intelligence testing by itself leads to the problems being placed on the child's head, instead of exploring whether the child's learning problem is created by the school. According to Aagaard, there is no contradiction between the use of psychometric methods and the investigation of pedagogical, cultural and societal causal relationships. By demonstrating normal abilities in students with poor learning outcomes, ability tests can help identify school-induced learning problems. Additionally, he argued, minority-language speakers can also have learning difficulties that are not caused by the minority-language situation, and psychometric tests can help to identify these. The problem is therefore not the testing itself, but that the test is too often used as a 'true measure of ability', due to the tester's lacking insight into how the test is constructed and into the effects of bilingualism.<sup>64</sup>

## Concluding Remarks

The transformation of the WISC-R from an American version into a Norwegian version was based on the premise that an individual's level of intelligence—the "abilities" that are measured by an intelligence test—is the product of both "nature" and "nurture". The test's concept of "ability" or "intelligence" was not distinct from the concept of achievement, and the test developers acknowledged that the test was culturally loaded. They also acknowledged that the learned abilities measured by the test were the kinds of abilities that aid success in an academically oriented school environment, and that

---

63 Aagaard, *Den språklige faktor*, p. 43, etc.

64 *Ibid.*, pp. 49–50.

these same abilities are trained through schooling. A schoolchild who scores well on a WISC-R test will, on average, do well in school; and doing well in school will increase their potential for doing well on a WISC-R test. The test score was, however, portrayed as something other than an ordinary school achievement test—it measured the result of the child's overall life experiences and learning both inside and outside of the classroom.

To transform the test into a Norwegian one, the test had to be culturally adapted. This cultural adaptation was not based on any explicitly defined concept of culture. However, it was implicitly based on the notion that Norwegian schoolchildren share a common and uniform culture. The test was not culturally adapted to specific groups within the total schoolchild population, and the question of social representativeness was mainly taken into consideration in the norming of the scale, and less in the cultural adaptation process. No specific measures were taken to ensure that the test items did not discriminate systematically between children with different life experiences due to different ethnic or socio-economic backgrounds. The transformation of the test seems, in contrast, to have been based on the expectation that IQ, and thus culture, was unevenly distributed between social strata. The test developers acknowledged that a WISC-R test score is a measure of learned and thus culturally loaded achievements. They also registered that the average working-class child achieved lower IQ scores than the average child of well-educated parents, and they saw this as a confirmation of the validity of the test, since it fit with similar findings in other Western nations. This means that the creation of the Norwegian version of the test was based on the expectation of low IQ scores among children with low socioeconomic status. Since an IQ score was considered a measure of abilities developed through culturally embedded life experiences, IQ was, in short, a measure of cultural conformity.

While the question of socioeconomic representativity was taken into consideration in the norming of the test, the question about cultural or linguistic differences was not taken into consideration at all. The transformation of the American test into a Norwegian one was based on the notion of an ethnically homogeneous nation. While the notion of a homogeneous nation was inaccurate in 1978, it became even less accurate during the test's life span of more than twenty years—a period

characterized by substantial immigration into Norway.

In addition to this, the test results, in Pihl's analysis, would often be interpreted as a measure of essential and given abilities—and not, as recommended by Kaufman, as a starting point for the development of hypotheses about the causalities behind the child's test results and IQ score. A score that, according to the *Handbook* and the test's originators, was a measure of how much the child had learned within a specific cultural context, here became interpreted as a measure of the child's inborn capacity for learning. The real-life outcome of this was a testing practice without racist intentions but with racist implications.

## Bibliography

- Aagaard, Knut Erik, *Den språklige faktor: Pedagogisk-psykologisk utredning av barn med minoritetsspråklig bakgrunn* (Doctoral dissertation, Oslo: University of Oslo, 2010).
- Diagnostiske hjelpemidler i den pedagogisk-psykologiske rådgivningstjenesten*, Innstilling avlevert november 1974 (Oslo, 1974).
- Egeberg, Espen, 'Bokomtale. Etnisk mangfold. Det sakkyndige blikket', *Spesialpedagogikk* 6 (2005), 42–43.
- Fancher, Raymond E., *The Intelligence Men: Makers of the IQ Controversy* (New York: WW Norton & Co., 1985).
- Fasting, Rolf, 'Kommentar til boka "Etnisk mangfold i skolen. Det sakkyndige blikket" av Joron Pihl', *Spesialpedagogikk*, 8 (2010), 47.
- Hagtvet, Knut A. and Johan O. Undheim, 'The Norwegian Experience of Test Use: A Selective Review of Norwegian Tests and Measurements in Cultural Context', in *Human Abilities in Cultural Context*, ed. by S. H. Irvine and J. W. Berry, 1st ed. (Cambridge: Cambridge University Press, 1988), pp. 263–81.
- Kaufman, Alan S., *Intelligent Testing with the WISC-III* (New York: Wiley, 1979).
- Kayser, Hans Georg, *Håndbok for tolkning av WISC-R* (Brandbu: Skolepsykologi - materiellservice, 1990).
- Kyllingstad, Jon Røyne, 'The absence of race in Norway?', *Journal of Anthropological Sciences, JASs Forum*, 95 (2017), 1–9. <https://doi.org/10.4436/JASS.95012>
- Mercer, Jane R., *Cultural Diversity, Mental Retardation, and Assessment: The Case for Nonlabeling*, Paper presented at the Fourth International Congress of the International Association for the Scientific Study of Mental Deficiency (Washington, D.C., 22–27 August 1976).

- Mønsterplan for grunnskolen* (Oslo: Kirke og undervisningsdepartementet/Aschehoug & Co., 1974).
- Nordahl, Thomas and Terje Overland, *Idealer og realiteter: Evaluering av spesialundervisningen i Oslo Kommune* (Oslo: NOVA Rapport 20, 1998).
- Nordland, Eva, *Innføring i psykologien* (Oslo: Aschehoug, 1955).
- Pihl, Joron, *Etnisk mangfold i skolen: Det sakkyndige blikket* (Oslo: Universitetsforlaget, 2010).
- Ramsøy, Natalie, 'Evner, utdanning og yrkesvalg i norsk samfunnsstruktur', *Tidsskrift for samfunnsforskning* (1961), 217–237.
- Statistical Yearbook of Norway* [Statistisk årbok] (Oslo: SSB, 1968–1978).
- Thrane, Vidkun Coucheron, *Avgangskarakterer i folkeskolen: Militærpsykologiske meddelelser no. k-1* (Oslo: Forsvarets psykologiske avdeling, 1961).
- Undheim, Johan Olav and David Wechsler, *Håndbok for Wechsler Intelligence Scale for Children—Revised*, Norwegian edition by Johan Olav Undheim (Oslo: Norsk psykologforening, 1978).
- Undheim, Johan Olav, 'On Intelligence IV: Toward a restoration of general intelligence', *Scandinavian Journal of Psychology* (1981), 251–265.
- Wechsler, David, *WISC-R: Wechsler Intelligence Scale for Children—Revised* (New York: The Psychological Corporation, 1974).
- Wesman, Alexander G., 'Intelligent testing', *The American Psychologist*, 23.4 (1968), 268–69.
- Westrheim, Karianne, 'Replikk', *Spesialpedagogikk*, 10 (2010), 32.
- Wold, Astri Heen, 'Kan for lite norsk, stemples som dum', *Tidsskrift for norsk psykologforening*, 1 December 2006.
- Williams, Robert L., *The BITCH-100: A Culture-Specific Test. Paper presented at the American Psychological Association Annual Convention* (Honolulu, Hawaii, September 1972).

## II. 'A Violation of the Child's Integrity and of Parental Rights': The 1959 Controversy on IQ Testing of Norwegian Schoolchildren

*Fredrik W. Thue*

---

In 1959, a heated debate broke out in the Norwegian press about mass intelligence testing of first graders in Oslo's elementary schools. Soon thereafter, a Member of Parliament raised the issue to the highest political level with an interpellation that led to an extensive debate in *Stortinget*, the Norwegian Parliament. The IQ dispute became the first in a series of related public and parliamentary controversies about the psychological and sociological surveying of schoolchildren. The debates played out in a similar manner; many of the same political and public actors participated, and opinion groups opposed each other in much the same way. The first dispute about IQ testing was also frequently invoked in the subsequent debates. The controversies had consequences—the use and administration of the intelligence tests, as well as the test itself, became substantially modified, and two other research projects on schoolchildren that had caused public outcry had to be terminated. These included the Norwegian part of the sociological Project Metropolitan, a large-scale comparative, longitudinal study on youth problems in the Nordic capitals.

The 1959 debate on mass intelligence testing of Oslo's school children presents a complex picture. As this chapter will show, the discussants mobilized an array of arguments with different cultural, political, and moral connotations. A centerpiece of the debate, however, was the claim that IQ testing of seven-year-olds and its use by teachers and school authorities violated the child's integrity and parental rights. The public



and parliamentary debate showed that this sensibility was widespread. It was also this argument that was recalled and revisited by MPs and other discussants in the later debates about research on schoolchildren, and which, in hindsight, appears as the unifying leitmotif of the controversies. This chapter will therefore pay special attention to this line of reasoning, while not ignoring the other arguments involved.

In what follows, I will first briefly describe the contested IQ testing of school starters in Oslo and its rationale as an aid to the progressive pedagogical transformation of the national school system. I will then untangle the different, closely intertwined dimensions of the IQ debate and the subsequent controversies of the scientific surveying of schoolchildren. In order to explain why such a seemingly limited issue as IQ testing of Oslo's school starters would raise national political concern, I will trace some inherent historical tensions within the Norwegian project of unified schooling, as well as some disagreements and conflicts of interests among the contemporary psychological and educational experts. Finally, I will discuss some possible implications for our understanding of the postwar Norwegian welfare state.

The chapter will serve two major purposes. By focusing on popular resistance to IQ testing in a Nordic country, the study fills a lacuna in the international historiography on intelligence testing. A strongly disproportional share of this research has focused on the United States, where intelligence testing obtained an extraordinary prevalence and impact. Even studies of criticism, controversies, and resistance to IQ testing have concentrated overwhelmingly on the US case and on the accusations in the 1960s and 1970s that intelligence testing reproduced and legitimized a starkly unequal, racist social order. As Annette Mülberger has pointed out, much of this historiography has strongly identified with such criticism and has tended to create a dystopian caricature of the history of intelligence testing.<sup>1</sup> More generally, John Carson has argued that the American historical experience with mental testing was 'anomalous' and that it needs to be 'provincialized' with the help of comparative historical research.<sup>2</sup>

---

1 Annette Mülberger, 'Biographies of a Scientific Subject: The Intelligence Test', in *Oxford Research Encyclopedias: Psychology* (Oxford: Oxford University Press, 2020).

2 John Carson, 'Mental Testing in the Early Twentieth Century: Internationalizing the Mental Testing Story', *History of Psychology*, 17 (2014), p. 254.

By bringing in a different kind of criticism from another period and national context, the study may be read as a modest corrective to this strongly US-centered research landscape. Unlike the later denunciation of IQ testing in the United States, which came mainly from the political left, the criticism raised in Norway was predominantly of a conservative or liberal-conservative nature—though not necessarily in a narrowly political sense. It was a conservatism better described as a general cultural and ethical sensibility shared by discussants from a relatively broad political spectrum.

Since mass intelligence testing of schoolchildren was justified in terms that related closely to the underlying logic of the upcoming welfare state, the controversy may also shed light on some of those tensions and conflicts that were inherent to that larger historical project. By exploring a case where the scientific and political logic of the welfare state collided with widespread, deep-seated norms and values that became sharply articulated, this chapter aligns with newer historiography that highlights the complexities and challenges of the Nordic social-democratic welfare state project—complexities that “Whiggish” interpretations of the welfare state as a consensual, progressive project have tended to underemphasize.

The controversies around intelligence testing in postwar Norway have been very sparsely researched and discussed. They have, however, been touched upon in a few studies within the Norwegian historiography on postwar education and educational science.<sup>3</sup> The later, related resistance against the ultimately cancelled Project Metropolitan has been subject to a *magister* thesis in sociology, as well as to several retrospective reflections by the project's prospective leader.<sup>4</sup> The most comprehensive scholarly

3 Alfred Oftedal Telhaug, *Den 9-årige skolen og differensieringsproblemet* (Oslo: Lærerstudentenes forlag, 1970); Kim G. Helsvig, *Pedagogikkens grenser. Kampen om norsk pedagogikk ved Pedagogisk forskningsinstitutt 1938–1980* (Oslo: Abstrakt forlag, 2005).

4 Ragnhild Hoëm, 'Om bakgrunn for reaksjonen mot Prosjekt Metropolitt, Department of Sociology, University of Oslo' (1968); Ørjar Øyen, 'Encounter with the Image of Sociology', *Sociologiske Meddelelser* 10 (1965), 47–60; Ørjar Øyen, 'Personvern og behovet for data til samfunnsvitenskapelig forskning', in *Personvern og samfunnsforskning* (= NSD Rapport nr. 17) (Bergen: Norsk samfunnsvitenskapelig datatjeneste 1977), pp. 3–15; Ørjar Øyen, 'Social Research and the Protection of Privacy: A Review of the Norwegian Development', *Acta Sociologica* 19 (1976), 249–262; Bente Gullveig Alver and Ørjar Øyen, *Forskningsetikk i forskerhverdag: vurderinger og praksis* (Oslo: Tano Aschehoug, 1997) pp. 76–101. Øyen was also Hoëm's

contribution of these controversies to date is Kim Helsvig's PhD thesis on the postwar controversies around the Department of Pedagogical Research at the University of Oslo. The resistance against IQ testing and Project Metropolitan is here approached primarily as part of a Christian-conservative opposition against the secular, naturalistic psychometric paradigm cultivated by the university department. However, Helsvig also points to the broad basis of the public outcry about the alleged violation of parental rights.<sup>5</sup>

All these interpretations, and especially Helsvig's thorough study, contain valuable insights. Still, I will argue that none of them do full justice to the complexity, prevalence, and depth of the motifs involved in the public resistance against IQ testing. While socially privileged parents and Christian-conservative critics certainly took an active part, the controversy engaged a much more diverse public, and testing was criticized from a wider range of political and socio-cultural standpoints than suggested by previous scholarship. A closer analysis and a wider perspective on the controversy is therefore warranted.

## The Functions of IQ Testing in the Oslo School: Segregation and Progressive Education

The intelligence testing of first graders in Oslo's elementary schools had been practiced for more than twenty years when the IQ dispute broke out in 1959. One day in October, class teachers administered an IQ test to their first graders. The test used was an adapted version of Kuhlmann-Anderson's group test from 1927, standardized on first graders in Oslo's primary schools in 1938–39 by Oslo's school inspector, Bernhof

---

supervisor, and the contributions of both clearly reflect his interests and involvement in the conflict. The public resistance against Project Metropolitan was, in their view, an expression of conservative suspicion towards the allegedly radical social sciences from upper-middle-class parents and their political advocates in the press and polity. These privileged groups also seemed to have a special aversion against having their "privacy invaded" by "inquisitive" psychologists and sociologists. A related interpretation of the resistance against intelligence testing of schoolchildren, which was hinted at both in the contemporary parliamentary debate and later by the educational historian Alfred Oftedal Telhaug, was that socially privileged parents feared the prospects of seeing their children being sorted out in the educational system based on their low IQ score. See Telhaug (1970), p. 76; St.forth. nr. 7b (1959), p. 2271 (statement by MP Rakel Seweriin).

5 Helsvig (2005), pp. 182–189.

Ribsskog, and the teacher Kristian Mønnesland.<sup>6</sup> After each pupil's IQ score had been calculated, the answers were sent to the municipal school inspector for control and statistics. The IQ was registered on the pupil's index card, and class teachers would also often make their own lists of the IQ scores, which they typically stored in a locked drawer in their desk in the classroom. Upon the pupils' final exam in the seventh grade, IQ scores from the first grade were transferred to the list for final assessment that were sent to the external examiners together with the examination papers.<sup>7</sup>

How was this testing practice justified, and what purposes did it serve?

Its first and most obvious purpose was segregation. From its very introduction into the educational system, intelligence testing had been employed to identify those "feeble-minded" or "mentally retarded" children who were not expected to profit from teaching in the unified school. While the group test applied to the entire population of first graders was different from those individual clinical tests that were used to diagnose the mentally weak, mass testing served as an initial screening that helped identify this sub-group. Children who scored among the lowest three percent would, with their parents' consent, be referred to Oslo's school psychology office, which would assess their need for special education with the help of individual tests and clinical observation.<sup>8</sup>

Interestingly, however, it was not primarily this most basic function of the group testing that was targeted in the public and parliamentary debate of 1959. Segregation of "mentally weak" children from the common school seems to have still been a widely accepted practice in the late 1950s. The legal scholar and criminologist Tove Stang Dahl has pointed to an intriguing dialectic in the development of the public Norwegian unified school system between integration and segregation: in order to persuade urban, upper-middle-class parents to send their

---

6 The Kuhlmann-Anderson group test succeeded Dearnborn's group test, which had been introduced in the Oslo school two years earlier. See Bernhof Ribsskog, *Kuhlmann-Andersons gruppeprøve. Standardisert for 1. klasse i folkeskolen. Rettledning for lærere* (Oslo: Gyldendal Norsk Forlag, 1947).

7 St.forh. nr. 7b (1959), p. 2265 (Borten); Emil Østlyngen, 'Gruppeprøver og deres bruk i Oslo folkeskole', *Nordisk psykologi*, 16 (1964), pp. 24–26.

8 St.forh. nr. 7b (1959) p. 2263 (Borten).

children to the public *folkeskole* (primary school) and not a private school, it was imperative to remove children who would otherwise disturb and hamper learning.<sup>9</sup> This segregation policy continued into and was even reinforced in the postwar period, until it was radically reversed in the 1970s.<sup>10</sup> That Oslo's test regime was applied to all schoolchildren and not merely to a selected outgroup of mentally weak or other deviants was probably an important reason why it provoked such strong public reactions.

Indeed, testing was partly justified in terms of the knowledge it produced about the great majority of children who would attend the unitary school, and who differed widely in their interests, abilities, and preconditions for learning. Since its introduction in Oslo's elementary schools in the late 1930s, group intelligence testing had been employed as an aid to pursue the focal objective of progressive, child-centered pedagogy—to adapt teaching and curriculum to the children's natural development and individual abilities.

In 1939, Norway had adopted a standard national curriculum for the seven-year mandatory *folkeskole* based on progressive, child-centered pedagogical principles, adapted to the requirements of a unified school system. This so-called *normalplan* [standard national curriculum] has, perhaps somewhat extravagantly, been described as the first scientifically based national curriculum in the world.<sup>11</sup> Intelligence testing and the normalization of grading played an important part in its development and implementation. Bernhof Ribsskog—a graduated educationalist, school inspector in the municipality of Oslo, and main architect of the *normalplan*—made a series of efforts in the 1930s to calibrate and normalize the grading scale of the *folkeskole*, modeled on the bell-shaped distribution of intelligence in the population. This normalizing effort went hand in hand with an appreciation of IQ testing as an aid to individualize teaching and evaluation. Gro Hanne Aas has argued that these seemingly contradictory utilizations of testing were reflective of

9 Tove Stang Dahl, *Barnevern og samfunnsvern: Om stat, vitenskap og profesjoner under barnevernets oppkomst i Norge* (Oslo: Pax forlag, 1978), p.120 ff.

10 Jan Froestad, 'Normalisering som disiplinering: Om norsk skolepolitikk', in *Normalitet og identitetsmakt i Norge*, ed. by Siri Meyer and Torvald Sirnes (Oslo: Ad Notam, 1997) pp. 76–98.

11 Gro Hanne Aas, 'Den nye skole: Enhetsskole, reformpedagogikk, karakterpolitikk', *Arr* 13 (2001), pp. 1–2: 25.

the efforts to implement a child-centered, progressive pedagogy within a unified national school system with a common curriculum. This was an unusual combination at the time. In other European countries, such as England, Germany, and Italy, varieties of the new pedagogy had predominantly been practiced in smaller, private, alternative schools.<sup>12</sup>

By the late 1950s, testing of schoolchildren was still legitimized as a technology supporting teachers in their efforts to adapt teaching to the abilities of each child. Defenders of testing frequently invoked a key sentence from the normal plan: 'The aim of education is not to help students with different abilities and potential become equally skilled in their schoolwork, but to give the students an education which as far as possible corresponds to their abilities and potential.'<sup>13</sup> Progressive educationalists and politicians hoped that intelligence tests would give a more objective, less socio-culturally biased measure of the child's inherent potential for learning and intellectual growth than teachers' more or less subjective judgment. Testing was thus appreciated by its defenders as an aid to the equalization of educational opportunities, as well as to the realization of progressive, child-centered forms of teaching.

In this way, testing was seen as instrumental to a unitary, democratic school system. The realization of such a system was a crucial dimension to the emerging Norwegian welfare state. By replacing an *institutionally* differentiated school system, with separate tracks for children from different social classes, with a comprehensive school that was *pedagogically* differentiated according to students' actual mental abilities, social equality would be reconciled with human diversity.<sup>14</sup>

In spite of broad general support of this educational policy, it was primarily the use of testing as an instrument of differentiation among "normal" children in the unified school that would raise concerns both in

12 Ibid.; Gro Hanne Aas, Bernhof Ribsskog, 'Elevenes læring i sentrum', in *Pedagogiske profiler: Norsk utdanningstenkning fra Holberg til Hernes*, ed. by Harald Thuen and Sveinung Vaage (Oslo: Abstrakt forlag, 2004), pp. 179–198.

13 E.g. Oslo's school inspector Eivind Jørgensen in his account for the test regime to the Municipal School Board, May 1959, quoted from 'Innstilling fra Testutvalget' 3 (Skoleetaten: Aa Møtebøker, Oslo skolestyre 6.11.1963, Oslo municipal archives), cf. Aas 2004: 191.

14 Telhaug 1970: 14 ff.; Florian Waldow, *Utbildningspolitik, ekonomi och internationella utbildningstrender i Sverige 1930–2000* (Stockholm: Stockholms universitets förlag, 2008), p. 92 f.

*Stortinget* and among the public. To understand why this specific aspect of a widely supported policy would provoke such fierce resistance, we must look more closely at the arguments and underlying emotions involved in the 1959 controversy.

### ‘A Very Dangerous, Slippery Slope’: The Parliamentary Debate of IQ Testing

The IQ controversy that broke out in the spring of 1959 was ignited by the voice of a well-known public person. In a nationally broadcast talk held as part of the programme series *The Parents’ Quarter*, Knut Tvedt, a Supreme Court lawyer well-known in Norwegian cultural life, targeted the testing of seven-year-old school starters. In a personal style, referring to his own memories from childhood, Tvedt described the IQ testing of pupils as an abuse of trust: of the children’s naïve, immediate trust in their teacher, as well as of the deep trust parents showed the school by confiding their beloved ones to its day-to-day custody.

A leitmotif was here introduced, which would reverberate widely in the debate that followed—that group IQ testing and similar forms of psychological and social inquiry constituted a violation of the child’s personal integrity as well as of parental rights. This was especially the case, Tvedt and several other discussants argued, when such tests and inquiries were authorized by the school board without the parents’ informed consent. This was the case with mandatory intelligence testing, as well as with several research projects that would be criticized in the press and from the rostrum of Parliament in the following years.

Per Borten, the MP who raised the issue in an interpellation to the Minister of Education two months later, characterized mandatory IQ testing as ‘an intervention in the psyche’ and an assault on ‘the individual’s integrity’. He warned that teachers, by overrating the value of an IQ test, could do an injustice to ‘the defenceless individual, the child’. The fact that the test results were kept secret from the children and their parents made IQ testing essentially different from ordinary school tests. This, he claimed, warranted the question: ‘To what extent was it legitimate to investigate the depths of a child’s soul—in secrecy?’<sup>15</sup>

---

15 St.forh. nr. 7b, 1959: 2265 (Borten).

Borten represented the Center Party, a non-socialist party with a core constituency of farmers and smallholders. His fellow party member Einar Hovdhaugen stated in the same debate: 'What is at stake here is the question of the integrity of the individual person. This is one of the most burning questions in our time, and in this case the question is all the more burning as its subject matter is the integrity of a child's mind.'<sup>16</sup> Sverre Stray from the Conservative Party warned that schoolchildren should not be used as 'scientific guineapigs' and pointed at the increasing scientific testing and surveying of pupils in school, which he feared would lead down 'a very dangerous, slippery slope'.<sup>17</sup>

Closely related to the concern about the child's personal integrity was the defence of parental rights and the integrity of the family as an institution. Several MPs criticized the fact that parents were largely kept in suspense about the tests and their child's IQ score, as well as about the way the results were used and stored. This half-secrecy vis-à-vis the home, it was argued, could undermine parents' basic trust in school—a trust that was essential for the child's upbringing and the school's social mandate.<sup>18</sup> Another important aspect regarded who was entitled to make decisions about the child's educational trajectory. Trygve Bull, a nonconformist representative of the Labor Party, and Olav Hordvik from the Liberal Party both expressed their worries that the schools' use of pupils' IQ scores would undermine the right of children and their parents to choose freely between educational alternatives after the unified *folkeskole*. They argued that parents and pupils should have an unrestricted right to decide freely, even if their decision might later prove wrong.<sup>19</sup>

In his answer to this broad criticism, Minister of Education Birger Bergersen assured his critics that qualified experts did not overrate

---

16 St.forh. nr. 7b, 1959: 2268.

17 St.forh. nr. 7b, 1959: 2274.

18 See, e.g., the statement of MP Berte Rognerud (Conservative Party), St. forh. nr. 7b, 1959: 2271.

19 St.forh. nr. 7b, 1959: 2268 f., 2272. This standpoint was adopted by a broad majority in Parliament two years later, when it discussed the basic principles of the new comprehensive school. The keynote of this debate was very similar to that of the IQ controversy. As the educational historian Alfred Oftedal Telhaug has summarized the session, one MP after the other stood up for democratic common sense and rejected the conclusions of the pedagogical experts and their ambition to control the school. Telhaug 1970: 73 f.



the predictive value of the tests and pointed to the heuristic value they had for the teacher when trying to facilitate learning and progress for each child. He also denied that IQ testing was essentially different from ordinary school tests. After quoting some very general and rather noncommittal sentences from the Ministry's proposed new instructions for teachers, he stated: 'As everyone will realize, guidelines have now been made that constitute a reassuring framework for different kinds of tests in school.' Bergersen concluded by pointing out that the scientific field in question was growing rapidly, and that 'it must under no circumstances be allowed that we, in our country, will do anything that may impede this development.'<sup>20</sup> An even stouter support of testing came from Labor MP Rakel Seweriin, who suggested that those parents who complained were simply disappointed with their children's scores.<sup>21</sup>

### IQ Testing and the Positivism Dispute: Philosophy Meets Democratic Common Sense

These answers hardly indicated that the Labor government was ready to back down on the use of group intelligence testing as a tool of pedagogical differentiation and school reform. Neither did it reveal much sensitivity towards those concerns that were raised by the critics. Perhaps the most eloquent and profound expression of the ethical sensibility that formed the basis of the widespread criticism of IQ testing was given by Per Lønning, a young MP from the Conservative Party. He pointed to the difference between an ordinary grade based on a pupil's achievements in the past, which he or she might improve in the future, and the IQ, which pretended to represent the individual and their potential future abilities. Lønning saw this as, in principle, an attack on human freedom: 'The individual is classified as a static, concluded entity. The human becomes an object, and there is no room for the living, dynamic personality'. He added that, 'this whole question about object and person has become subject to a very comprehensive analysis within contemporary philosophy'. IQ testing therefore raised questions that were closely related to human dignity and human rights.<sup>22</sup>

---

20 St.forh. nr. 7b, 1959: 2265.

21 St.forh. nr. 7b, 1959: 2271.

22 St.forh. nr. 7b, 1959: 2266 f.

Lønning was a theological scholar, a pastor, and an active member of the conservative student group Minerva, which cultivated a "pre-political" style of thought. Oriented towards the fundamental cultural and ethical preconditions of democracy and the rule of justice, Minerva represented a distinctive brand of conservatism, inspired by existentialism as well as neo-Thomism and continental Christian Democracy.<sup>23</sup>

Lønning therefore had thorough knowledge of the philosophical critique of positivism, which would become very influential in parts of Norwegian intellectual life in the 1960s and 1970s. A year before the IQ controversy broke out, the young philosopher Hans Skjervheim had published an essay, titled 'Participant and Observer', which would later figure prominently in the national anti-positivist canon. Here, he pointed to the ethical problem of reducing the other to an object in one's own world, as psychologists and social scientists allegedly did when they studied humans. That people tended to feel uncomfortable and to defend themselves against the penetrating glance of the psychologist, Skjervheim argued, was actually a perfectly reasonable response to his objectivizing attitude, for this epistemic attitude entailed in principle an attack on the other's freedom.<sup>24</sup>

Lønning's statement in the IQ debate largely reiterated this argument. As probably the only philosophically educated MP at the time, he was uniquely equipped to frame the issue in this manner. What is interesting to note, however, is the extent to which other MPs, most of whom had little or no academic education, expressed much of the same ethical sensibility, and often in quite a pertinent way. Hence, to the extent that the emerging philosophical critique of positivism reverberated with the IQ controversy, it was probably not due to academic philosophy's impact on politics and public discourse, but rather the other way around: Skjervheim and other critics articulated in a philosophical language an ethical awareness that was already in the air among reflective ordinary people.

23 Johannes W. Løvhaug, *Politikk som idékamp. Et intellektuelt gruppeportrett av Minervakretsen 1957–1972* (Oslo: Pax forlag, 2007).

24 Hans Skjervheim, 'Deltakar og tilskodar (1957)', in *Skjervheim, Deltakar og tilskodar og andre essays* (Oslo: Tanum-Norli, 1976) p. 56.

## Expert Criticism of Oslo's Testing Practice: Conflicting Professional Interests?

The early postwar period has conventionally been portrayed as a high tide of popular trust in research and scientific expertise in the Western world. This has also been strongly emphasized in much historiography on the Nordic welfare states. While the IQ debate may tell us something interesting about the limits of such trust, the political targeting of testing was also facilitated by experts' public criticism of the way Oslo's schoolchildren were tested and the manner in which results were stored and utilized. Several prominent psychologists took part in the debate and underscored the limited reliability and validity of the intelligence quotient. What came across in these debates was that scholarly beliefs that were said to have initially formed the basis of the IQ—that intelligence was an inherited, biologically determined quality which remained relatively stable throughout one's lifetime—had largely been refuted. Intelligence was now understood as a much more dynamic quality, affected by the individual's environment, education, and life course.<sup>25</sup> Both psychologists and educationalists criticized the narrow focus on intelligence and turned towards a more comprehensive study of the individual's personality, of which intelligence was only a part. While such criticism was almost as old as the IQ itself, and had been discernible among Norwegian psychologists and "psycho-technicians" at least from the early 1930s, it was now presented by psychologists and educationalists as scholarly common sense.<sup>26</sup>

The experts' eagerness to criticise group intelligence testing administered by teachers also related to their vested interests. The psychologists were particularly eager to emphasize the indispensability of clinical observation, individual testing, and educated professional judgment when assessing a child's mental abilities. In an article titled

25 'Ukritisk gruppetesting av 1. klassinger meningsløs og farlig. Testene er hjelpemiddel for vitenskapen, et barns I.Q. skal være ukjent for lærerne, sier professor Harald Schjelderup', *Dagbladet*, 14 December 1959.

26 Cf. Isak Lønne Emberland, *Den omhyggelig planlagte livsbane: Psykoteknikken og profesjonaliseringen av norske psykologer 1920–1947* (Ph.D. thesis, Oslo: Centre for the Study of Professions, Oslo Metropolitan University, 2024), p. 219 f. Professor Sandven's psychometric approach within education also aimed at measuring a wider spectrum of personality traits than mere intelligence; see Helsvig 2004: 58–85.

'Away from the unfortunate IQ' in the liberal daily *Dagbladet*, Arvid Ås, assistant professor in psychology at the University of Oslo, denounced the use of group tests in Oslo's elementary schools as irresponsible from a scientific as well as from an ethical point of view.<sup>27</sup> His criticism was fully supported by his senior colleague, professor Harald Schjelderup, who explained that an IQ test taken at the age of seven had very little predictive value; only from the ages of fifteen to seventeen would such a test be able to forecast the pupil's future intelligence level. Schjelderup was emphatic that teachers should not have access to their pupils' IQ scores and condemned the 'uncritical group testing of first graders' as 'meaningless and dangerous'.<sup>28</sup> Further, professor of pedagogy Johs Sandven, an ardent advocate of the psychometric approach in education, acknowledged in an interview with the conservative *Aftenposten* that 'we should possibly avoid the concept intelligence quotient in certain contexts', since it was so easily misunderstood and reified by the lay public.<sup>29</sup>

Critical parents, politicians, and commentators could thus find substantial support of their concerns among scholars. Indeed, the first to blow the whistle about the risk of misinterpretation and misuse of pupils' IQ scores, Oddvar Vormeland, was himself an educationalist and senior employee at Oslo's child guidance office.<sup>30</sup> He had, in late 1957, written to the municipal school inspector and warned that IQ scores were stored and treated in a manner that was less than reassuring. They were often easily accessible, even to short-term substitute teachers, and most teachers lacked the competence to assess them with necessary caution. Vormeland had felt obliged to bring this issue to the attention of the school inspector 'after having seen too many examples of how the test scores could be misused—and are misused'.<sup>31</sup> This first warning,

27 'Bort fra den ulykksalige I.Q. Både faglig og etisk uforsvarlig å bruke gruppeprøver slik som det gjøres i folkeskolen', *Dagbladet*, 18.04.1959.

28 'Ukritisk gruppetesting av 1. klassinger meningsløs og farlig. Testene er hjelpemiddel for vitenskapen, et barns I.Q. skal være ukjent for lærerne, sier professor Harald Schjelderup', *Dagbladet*, 14 December 1959.

29 'Vi bør muligens forlate begrepet intelligenskvotient', *Aftenposten*, 31 October 1959 (evening).

30 Vormeland is an interesting intermediary actor in this conflict: While educated within Sandven's psychometric paradigm at the department of pedagogical research, he was also an active member of the anti-progressivist Institute for Christian Upbringing and editorial secretary of its journal *Prismet*.

31 Quoted from 'Innstilling fra Testutvalget', p. 14 (Skoleetaten: Aa Møtebøker, Oslo

however, did not lead to substantial modifications of the test regime. Although the school inspector in Oslo largely supported Vormeland's views, the capital's collegium of school leaders decided to continue the established practice, including the registration of IQ on the pupils' index cards and the transfer of this information to the list for final assessment at the end of the seventh grade.<sup>32</sup> Teachers and school leaders thus seemed to be more eager than both psychologists and educationalists to continue testing and to use the scores as a way for teachers to assess the children's mental abilities and future intellectual capacity. Those who blamed the psychological and educational scientists for the misuse of IQ testing in the Oslo school may therefore have barked up the wrong tree. In the following few years, however, social scientists would become more directly targeted in the public and parliamentary protests against scholarly surveys of schoolchildren.

### Integrity vs. the Quest for Knowledge: Political Demands for a Social Research Ethics

Skjervheim's criticism of the objectivist epistemic attitude of psychologists and other social scientists emphasised the integrity of the person who was being surveyed. The slightly different issue of parental rights, while an important element of the IQ dispute, would become the dominant concern in two subsequent controversies related to the social-scientific surveying of schoolchildren. In late 1959, a heated debate arose in the press after a student of pedagogy had distributed an extensive survey to pupils in Oslo's lower-secondary schools as part of her master's thesis on juvenile delinquency and dropouts from school. The project had been approved by the Department of Pedagogical Research and Oslo's municipal school board, but no information had been given to the parents. This procedure, as well as the character of the questions, which partly probed into private issues such as the respondents' relations to their family, friends, and school, stirred outrage among parents and sharp criticism in the conservative press. When the case was brought up in Parliament, neglect of parental rights and violation of the pupils' and

---

skolestyre 6 November 1963, Oslo municipal archives).

32 Ibid.: 17 f.

their families' integrity became the two closely related focal points. The invitation to evaluate one's family and reveal details of private family life without parental consent was seen by some discussants as in itself indecent and corrosive not only of public trust in social research, but also of the personal, spontaneous trust between family members.<sup>33</sup>

Such criticism was raised in an even sharper form four years later, when a team of Nordic sociologists launched a comparative, longitudinal, in-depth study of the coming of age of a selected group of males of the 1953 cohort in the capitals Helsinki, Stockholm, Copenhagen, and Oslo. The project would monitor its research subjects from boy- to adulthood and extract a wide range of data on their family, friends, educational and occupational trajectories, possible criminal record, etc. It was thus much more invasive than the project that caused the previous controversy, with more lasting repercussions on the lives of the research subjects and their social environment. Scholars and school authorities had apparently not learnt from previous experience; the initial questionnaires were yet again approved by Oslo's school board and distributed to pupils without their parents' knowledge. The scholars' assurance that nobody would be enlisted in the longitudinal project without their own and their parents' consent did not save the project from massive public criticism. The conservative MP Per Lønning, who brought up the issue in Parliament, concluded that the time was now ripe to introduce a set of permanent rules for how parents should be involved in such cases. The purpose was not only to protect the integrity of private life and the home, but also to restore the citizens' trust in the school, as well as in scientific research.<sup>34</sup>

Another representative of the conservative party, Berte Rognerud, called it a fact that social research was being watched with considerable suspicion and even outright distrust by the public. Part of the reason, she reflected, was that the social sciences were expanding at the same time as state control and intervention was taking effect in each person's daily life and behavior in nearly all areas. The public tended to mix up these two trends and blame the social sciences for statist expansionism, which, she admitted, might not be entirely fair. But she also found reason to proclaim that 'the children belong to their parents, not to society'.<sup>35</sup>

---

33 St.forh. 1959/60, nr. 7a: 1096 (Rognerud), 1100 (O. Lyng).

34 St. forh. 1964/65, nr. 7b: 2071–2072

35 St. forh. 1964/65, nr. 7b: 2075–2076.

Lønning, in a similar vein, pointed to Orwell's and Huxley's dystopian visions and to the abuse of children that took place by involving them in the political surveillance of their parents in East Germany.<sup>36</sup>

How did the Labor government respond to this increasingly sharp criticism? A recurring argument was that one should trust those psychologists and social scientists whose research findings and expertise could contribute to the solution of urgent social problems. In the controversy of late 1959, Minister of Education Birger Bergersen argued that social-scientific inquiries would, in most cases, lose their value if parents and/or pupils were informed about their content in advance. Public trust was therefore vitally important. He concluded that pedagogical research must be provided with 'the best and freest working conditions possible'.<sup>37</sup> In 1965, his successor Helge Sivertsen would likewise summon the parliamentarians to educate the public, rather than cater to its views: 'If *Stortinget* supports social science with its full strength, it will, no doubt, have a great impact on public opinion.'<sup>38</sup> His view was strongly supported by Labor's Rakel Seweriin, who consistently argued that the social sciences encountered conservative prejudice and that the public was in need of enlightenment.<sup>39</sup>

And yet, in the 1965 Project Metropolitan debate, there was a perceptible change of tone from the Labor government. Minister Sivertsen now emphasized that the expansive social sciences needed an elaborated research ethics, similar to the ethics that had earlier been developed in medicine and was underway in psychology.<sup>40</sup> He announced that the University of Oslo had appointed a committee to discuss principles for

36 St. forh. 1964/65, nr. 7b: 2075.

37 St. forh. 1959/60, nr. 7a: 1095.

38 St. forh. 1964/65, nr. 7b: 2081.

39 St. forh. 1959/60, nr. 7a: 1105.

40 The formation of a code of research ethics does not seem to have proceeded very far within Norwegian psychology, however. At the fourth meeting of the Nordic Association of Psychologists in 1956, Einar Dannevig stated that Norwegian psychologists had not yet 'arrived at a common presentation or statement of those questions that concern our occupational ethics'. This, he suggested, was partly due to strong differences of opinion on these issues within the profession. He also argued that there was widespread scepticism about committing to a definite set of rules before having deliberated their practical implications carefully. (*Fjerde Nordiske Psykologmøde i København 1956* (= Tillæg til *Nordisk Psykologi*) (København: Ejnar Munksgaards forlag, 1957), p. 53. A brief discussion about research ethics in the student psychology journal *Impuls* of 1964 hardly indicates that the profession had come very much further at that point in time.

social researchers to access public archives and registers, and ethical guidelines for the handling of such data.<sup>41</sup> The controversy that erupted over Project Metropolitan has later been portrayed as a major catalyst for the elaboration of social research ethics in Norway.<sup>42</sup>

The way public and parliamentary debate developed between the IQ controversy of 1959 and the Project Metropolitan scandal of 1964–65 thus became a turning point that forced both public authorities and the social-scientific community to pay more attention to the conflict between the quest for psychosocial knowledge and citizens' right to personal integrity. In late 1959, Oslo's municipal school board appointed a committee with scientific experts and school people, as well as two of the most profiled public opponents of the IQ test regime to critically examine current testing practices and propose possible changes. The committee's report led to substantial changes: testing was continued, but the test was revised; parents would be better informed; and the results were no longer communicated to the school in the form of a fixed IQ score, but instead roughly indicated by the categories 'critically weak', 'average', or 'substantially beyond average'. Perhaps most importantly, test results would be deleted after the first term, and would no longer be registered on the pupil's index card or on the final assessment lists at the seventh grade.<sup>43</sup> The two subsequent controversies had an even more evident impact—both the detailed psychological-pedagogical inquiry of lower-secondary school children and the Norwegian part of Project Metropolitan were eventually abandoned.<sup>44</sup> Hence, while faith, particularly on the left, in progressive social reform supported by social research was by no means dead, widespread public skepticism had now placed some limits around the project.

Was this merely a temporary setback from the ongoing development of a scientifically supported welfare state? In one sense, the answer is obviously yes—the social sciences underwent tremendous expansion from the early 1960s, and have long become an integral and vital resource in the administration of the Norwegian welfare state. To the extent that the protests against mandatory IQ testing and invasive surveys of

---

41 St. forh. 1964/65, nr. 7b: 2073.

42 Alver and Øyen 1997: 85.

43 Østlyngen 1964; Innstilling fra Testutvalget (Skoleetaten: Aa Møtebøker, Oslo skolestyre 6.11.1963, Oslo municipal archives).

44 Hoëm 1968: 49.



schoolchildren targeted the social sciences as such, they did not achieve their objective. However, while some of the public discussants and one or two MPs voiced a general dislike of the social sciences, almost all the parliamentary critics were careful to underscore that they did not want to hamper the development of psychological, pedagogical, or sociological research; their concern was to secure public trust in social research by assuring that it was performed in a trustworthy manner that respected individual integrity and parental rights.

### The IQ Dispute: A Symptom of Tensions in National Policies of Unified Schooling

In order to understand why such a highly specific and seemingly limited issue as the practice and use of IQ testing in Oslo's elementary schools would ignite such extensive and heated debate in 1959, it could be fruitful to construe it as a condensed expression of more enduring conflicts and tensions inherent to the national tradition of unified schooling. Such tensions were brought to the fore in the 1950s, as the governing Labor Party charted the course for the introduction of nine years of comprehensive schooling for children in a strategic manoeuvre that has been aptly characterized as 'de-politization under strong political control'.<sup>45</sup> This educational policy and implementation strategy largely followed the example of neighboring Sweden. The Labor parties of both countries won landslide victories in election after election in the postwar period and effectively controlled the national political process by means of a clear majority in Parliament, backed by strong party discipline. However, instead of seeking immediate parliamentary support for extended mandatory schooling, the Swedish Labor government had introduced a law in 1950 on experimental school development, which seemingly left it to the educational experts to decide how a nine-year comprehensive school could be optimally designed and implemented.<sup>46</sup> Four years later, the Norwegian Parliament passed a very similar law on

45 Rune Slagstad, 'Arbeiderpartistaten som skolestat', *Nytt Norsk Tidsskrift*, 11 (1994), 3–4: 226.

46 Gunnar Richardson, *Drömmen om en ny skola. Idéer och realiteter i svensk skolpolitik 1945–1950* (Stockholm: Liber 1983); Christian Lundahl, 'Progressiva koalitioner, (inter)nationella influenser och kunskapsmätningar i reformarbetet med svensk läroplan, ca. 1930–1950', *Nordic Journal of Educational History*, 1 (2014), 1: 70 f.

innovations and experiments in the school system, which provided the legal basis of a new National Council for Innovation in Education. In effect, Labor politicians and a selected network of experts and strategists within the educational field formed a strong alliance, leaving the non-socialist opposition virtually out of the political process.<sup>47</sup>

This situation brought to the fore some deep-seated tensions in the national project of unified education. While largely copying Sweden's nine-year comprehensive school scheme after World War Two, Norway had introduced a unified elementary-school system well before its eastern neighbor, and had its own distinctive tradition of democratic schooling. A public *folkeskole*, which was intended to be common for all children until the fifth grade, was legislated by a liberal-progressive government in 1889, and in 1920, seven years of common schooling was effectively implemented. But this national project of unified, democratic education came with some inherent ambiguities. The *folkeskole* that had been introduced in 1889 in the wake of Norway's parliamentary "revolution" of 1884 was envisioned and designed as a school run and controlled by the local municipality. The governing liberal-progressive party saw the school as an aid to and an extension of the parental upbringing of children—since the school, in principle, belonged to the parents, it should be controlled by the local community and the common sense of ordinary men and women.<sup>48</sup> But the liberal-democratic movement also saw the unified *folkeskole* as an instrument of national integration and the levelling out of class differences in society. Thus, the school appeared both as based on a parental mandate and as the strategic vehicle of a large-scale transformation of society.<sup>49</sup> While this latter, macro-reformist perspective would become gradually more dominant throughout the inter-war and early postwar period, the former, localist-parental understanding of the school's mandate did not disappear. During the years of German occupation from 1940–1945, as the collaborationist regime attempted to remake the national school system into an instrument of Nazification, this view of the school's mandate and responsibility was mobilized as the normative basis of

47 Alfred Oftedal Telhaug, *Forsøksrådet for skoleverket 1954–1984. En studie i norsk skoleutvikling* (Oslo: Universitetsforlaget, 1990) p. 25 ff.

48 Hans- Jørgen Dokka, *Fra allmueskole til folkeskole: studier i den norske folkeskoles historie i det 19. århundret* (Oslo: Universitetsforlaget, 1967) p. 407.

49 *Ibid.*: 409.

organized resistance. The widespread claim in the 1959 controversy that intelligence testing violated parental rights could be seen as a late manifestation of the tenacity of this dimension of Norway's democratic school heritage.

Another conflict within the Norwegian tradition of elementary schooling was the divide between the rural and the urban *folkeskole*. This schism, rooted in the very origins of the Danish-Norwegian *allmueskole* (commoners' school) in the eighteenth century, had proved strikingly persistent throughout the history of Norwegian education. By 1959, the rural school system had been significantly modernized and centralized, but children living in urban and suburban areas still received substantially more schooling than children from the countryside. While the rural school had been dominated by male teachers of peasant background with a Christian and nationalist vocational ethos, the urban school had a much larger element of female teachers of urban, middle-class origins, who tended to be more inclined towards modern educational science and child-centered pedagogical methods than their male colleagues in the rural school.

The Oslo school had long enjoyed a special status nationally as a vanguard and showcase of progressive pedagogy and educational reform. The government's postwar educational policies were very much based on the paradigm of the urban school and particularly on the achievements in Oslo. The close alliance that was established between Labor politicians and educational expertise in the 1950s had a precedent in Oslo, where the educationalist and Labor sympathizer Bernhof Ribsskog had combined research, local experiments, and national curricular reform in his capacity as municipal school inspector from 1929 to 1953.<sup>50</sup> The great importance that Labor ascribed to school development in Oslo was reflected in the city's politically composed school board. From 1946 to 1962, it was headed by the secretary of the Prime Minister's office. By the late 1950s, the Prime Minister and Minister of Education could also easily monitor its negotiations through their wives, who both happened to be members of the board. The board also included the editor-in-chief of Labor's central mouthpiece *Arbeiderbladet*, a crucial position of power within the party at that time.

This extraordinary mustering of political heavyweights on Oslo's

---

50 Aas 2001; Aas 2004.

school board could be seen as reflective of a more general tendency within the postwar Labor Party. As it seemed destined to be in office on a virtually permanent basis, it concentrated its leadership elite in the capital to an extent that broke markedly with the traditional ethos of representative democracy in Norway.<sup>51</sup> This concentration of executive political power, a source of increasing frustration and a sense of powerlessness within the parliamentary opposition, was not least striking within the field of education and the ongoing reform of the school system—a policy area where the values and political judgment of laypeople and local communities had previously been deemed as particularly relevant.

The IQ dispute of 1959 could at least partly be seen as a thinly veiled protest against these closely interrelated trends towards centralism, technocracy, and Labor omnipotence within national educational politics. This would help explain why the criticism of intelligence testing in Norway in the 1950s was launched mainly from the right and was predominantly culturally conservative in character, in contrast to the North American criticism of the IQ in the 1960s and 1970s, which came from the political left and highlighted the unjust social and racist effects of testing.

### Competing Postwar Political Ethics: Statist Individualism vs. Inviolable Privacy

The postwar efforts to equalize educational opportunities, and replace a differentiated school system with a unified school where teaching was adapted to individual abilities offers a striking example of how the welfare state attempted to emancipate the individual from their dependency on family, geography, and social class. The Swedish historian Lars Trägårdh has coined the term 'statist individualism' to capture this crucial dimension of the Nordic welfare state project. A new "social contract" would help liberate the individual citizen from their dependency on family, fortune, and paternalistic forms of benevolence by means of social rights and benefits provided by the

---

51 Cf. Henry Valen, *Nominasjon av stortingskandidater i Det norske arbeiderparti* (Master's thesis, Oslo: University of Oslo, 1954).

state on a strictly individual basis.<sup>52</sup> The concept of statist individualism dovetails neatly with the intentions behind intelligence testing. As it was constructed, applied, circulated, and received throughout the Western world and beyond, IQ testing promised to reveal each person's innate mental abilities and potential for intellectual development, relatively independent of his or her social background and cultural capital. The introduction and spread in the interwar and early postwar era of mass intelligence testing as a substitute for or supplement to established ways of ordering and selecting people could be seen as part of this process of state-supported individualization. But mass intelligence testing may also be construed as part of a modern "bio-politics" aimed at mobilizing and utilizing the nation's totality of "human resources" in a maximally efficient manner.

The pattern of protest and controversy I have traced in this chapter, however, hardly indicates that statist individualism was a widely accepted social contract in postwar Norway. While offering a pertinent characterisation of the broadly popular educational policies of Norwegian (and Swedish) Labor governments, statist individualism seems in this case to have collided with widespread attitudes and values in the Norwegian population.

The arguments applied by the critics of mass intelligence testing and other mandatory surveys of schoolchildren were more or less in line with the UN's postwar declaration of universal human rights, which had established the inviolability of personal integrity as a basic ethical and legal principle. This inviolability also applied to the family; article 16 claimed that 'the family is the natural and fundamental group unit of society and is entitled to protection by society and the State'. This principle did not sit well with statist individualism, and the IQ controversy could be read partly as a tug-of-war between the ethos of the human rights declaration and that of welfare state interventionism.

Although it cannot be definitely proved, I would hypothesize that one of the reasons why the conservative arguments in the IQ controversy reverberated so widely in Parliament as well as in public opinion was that they implicitly recalled the crucial, symbolically loaded memories of national civilian resistance during the years of German occupation.

---

52 Lars Trägårdh, 'Det civila samhällets karriär som vetenskapligt och politiskt begrepp', *Tidsskrift för samfunnsforskning*, 49 (2008), 4: 588.

The core argument that had been mobilized in the successful resistance against the enforced Nazification of schooling and upbringing—a resistance that had united teachers, clergy, and parents in an alliance that had proved invincible—was that the school and its teachers were primarily accountable to the children, their parents, and the Norwegian people. Only secondarily were they accountable to the state, and only insofar as it respected this primary source of legitimacy.<sup>53</sup> While none of the discussants in the controversies I have analyzed explicitly invoked the wartime experience, scattered references to Orwell, Huxley, and the German Democratic Republic indicate that fear of a totalitarian slippery slope was an important motive for some of the critics.<sup>54</sup>

To what extent was this a passing sentiment, aroused by a specific discursive situation? To what extent have we come across certain cultural attitudes and political norms that played a more crucial part in the negotiations that shaped the postwar Norwegian welfare state than the dominant, consensual historical narrative has acknowledged? These questions can only be answered through historical research of a much wider scope than this minor case study.

## Bibliography

Alver, Bente Gullveig and Ørjar Øyen, *Forskningsetikk i forskerhverdag: vurderinger og praksis* (Oslo: Tano Aschehoug, 1997).

Carson, John, 'Mental Testing in the Early Twentieth Century: Internationalizing the Mental Testing Story', *History of Psychology*, 17 (2014), 3: 249–255, <https://doi.org/10.1037/a0037475>

Dahl, Tove Stang, *Barnevern og samfunnsvern: Om stat, vitenskap og profesjoner under barnevernets oppkomst i Norge* (Oslo: Pax forlag, 1978).

53 See e.g. Nicola Kristin Karcher, *Kampen om skolen: nazifisering og lærernes motstand i det okkuperte Norge* (Oslo: Dreyer 2018).

54 However, in a debate in Parliament the same spring on the future organization of the lower- and upper-secondary schools, Lønning explicitly invoked the wartime experience. In a polemic against Labour's Rakel Seweriin, who denounced private schools as a threat to the unified school system with its alleged tolerance and pluralism, Lønning remarked shrilly: 'It is primarily each home that is responsible for the children's upbringing. This was what we fought for during the war, when so-called "societal concerns" demanded that the young should not be brought up based on parents' so-called "limited point of view".' St. forh. Nr. 7b, 1959: 1774, quoted from Vegard Kvam, *Opplæringsfrihet: pedagogiske alternativer til den offentlige fellesskolen*, Oslo, Universitetsforlaget 2024: 105.

- Dokka, Hans- Jørgen, *Fra allmueskole til folkeskole: studier i den norske folkeskoles historie i det 19. århundret* (Oslo: Universitetsforlaget, 1967).
- Emberland, Isak Lønne, *Den omhyggelig planlagte livsbane: Psykotenikken og profesjonaliseringen av norske psykologer 1920–1947* (Ph.D. thesis, Oslo: Centre for the Study of Professions, Oslo Metropolitan University, 2024).
- Fjerde Nordiske Psykologmøde i København 1956 (= Tillæg til Nordisk Psykologi) (København: Ejnar Munksgaards forlag, 1957).
- Froestad, Jan, 'Normalisering som disiplinering: Om norsk skolepolitikk', in *Normalitet og identitetsmakt i Norge*, ed. by S. Meyer and T. Sirnes (Oslo: Ad Notam, 1997) pp. 76–98.
- Helsvig, Kim G., *Pedagogikkens grenser. Kampen om norsk pedagogikk ved Pedagogisk forskningsinstitutt 1938–1980* (Oslo: Abstrakt forlag, 2005).
- Hoëm, Ragnhild, 'Om bakgrunn for reaksjonen mot Prosjekt Metropolit' (Master's thesis, Oslo: Department of Sociology, University of Oslo, 1968)
- Karcher, Nicola Kristin, *Kampen om skolen: nazifisering og lærernes motstand i det okkuperte Norge* (Oslo: Dreyer, 2018).
- Kvam, Vegard, *Opplæringsfrihet: pedagogiske alternativer til den offentlige fellesskolen* (Oslo: Universitetsforlaget, 2024).
- Løvhaug, Johannes W., *Politikk som idékamp. Et intellektuelt gruppeportrett av Minerva-kretsen 1957–1972* (Oslo: Pax forlag, 2007).
- Lundahl, Christian, 'Progressiva koalitioner, (inter)nationella influenser och kunskapsmätningar i reformarbetet med svensk läroplan, ca. 1930–1950', *Nordic Journal of Educational History*, 1 (2014), 1: 59–79.
- Mülberger, Anette, 'Biographies of a Scientific Subject: The Intelligence Test', in *Oxford Research Encyclopedias: Psychology* (Oxford: Oxford University Press, 2020), <https://doi.org/10.1093/acrefore/9780190236557.013.694>
- Ribsskog, Bernhof, 'Kuhlmann-Andersons gruppeprøve. Standardisert for 1. klasse i folkeskolen. Rettleiing for lærere' (Oslo: Gyldendal Norsk Forlag, 1947).
- Richardson, Gunnar, *Drømmen om en ny skola. Idéer och realiteter i svensk skolpolitik 1945–1950* (Stockholm: Liber, 1983).
- Skjervheim, Hans, 'Deltakar og tilskodar (1957)', in *Deltakar og tilskodar og andre essays*, ed. by H. Skjervheim (Oslo: Tanum-Norli, 1976), pp. 51–72.
- Slagstad, Rune, 'Arbeiderpartistaten som skolestat', *Nytt Norsk Tidsskrift*, 11 (1994), 3–4: 221–233.
- Telhaug, Alfred Oftedal, *Den 9-årige skolen og differensieringsproblemet* (Oslo: Lærerstudentenes forlag, 1970).
- Telhaug, Alfred Oftedal, *Forsøksrådet for skoleverket 1954–1984. En studie i norsk skoleutvikling* (Oslo: Universitetsforlaget, 1990).

- Trägårdh, Lars, 'Det civila samhällets karriär som vetenskapligt och politiskt begrepp', *Tidsskrift för samhällsforskning*, 49 (2008), 4: 575–594.
- Valen, Henry, *Nominasjon av stortingskandidater i Det norske arbeiderparti* (Master's thesis, Oslo: University of Oslo, 1954).
- Waldow, Florian, *Utbildningspolitik, ekonomi och internationella utbildningstrender i Sverige 1930–2000* (Stockholm: Stockholms universitets förlag, 2008).
- Østlyngen, Emil, 'Gruppeprøver og deres bruk i Oslo folkeskole', *Nordisk psykologi*, 16 (1964), 24–26.
- Øyen, Ørjar, 'Encounter with the Image of Sociology', *Sociologiske Meddelelser*, 10 (1965), 47–60.
- Øyen, Ørjar, 'Personvern og behovet for data til samfunnsvitenskapelig forskning', in *Personvern og samfunnsforskning*, NSD Rapport nr. 17 (Bergen: Norsk samfunnsvitenskapelig datatjeneste, 1977), pp. 3–15.
- Øyen, Ørjar, 'Social Research and the Protection of Privacy: A Review of the Norwegian Development', *Acta Sociologica*, 19 (1976), 3: 249–262.
- Aas, Gro Hanne, 'Bernhof Ribsskog: Elevenes læring i sentrum', in *Pedagogiske profiler: Norsk utdanningstenkning fra Holberg til Hernes*, ed. by H. Thuen and S. Vaage (Oslo: Abstrakt forlag, 2004), pp. 179–198.
- Aas, Gro Hanne, 'Den nye skole: Enhetsskole, reformpedagogikk, karakterpolitikk', *Arr*, 13 (2001), 1–2: 22–31.





## 12. Tests, Metrics, and the Making of the Norwegian School, 1950–2025

*Kim Helsvig*

---

During the 1950s, a cooperation between research interests and political interests was established in Norway on the subject of psychometric testing for selection and differentiation within the emerging comprehensive social democratic elementary school. From the late 1960s, however—in the context of the international *Positivismusstreit* in the social sciences—both the dominant Labor Party and the Ministry of Education and Church Affairs dismissed the psychometric tradition as an irrelevant policy partner for developing the school based on the notions of equity, equal opportunities, and national and social integration.<sup>1</sup> From the turn of the century, the “PISA era” has nevertheless seen a huge revival of educational tests and measurements as part of a decisive and OECD-inspired international turn, also reflected in Norwegian educational debate and policy.<sup>2</sup> This chapter addresses four closely related research questions: (1) Why and how were new psychometric tests introduced in Norwegian schools in the 1950s? (2) Why were they abandoned for almost three decades from the 1970s? (3) Why and how were psychometric tests and tools reintroduced on a large scale after the publication of the first Program for International Student Assessment study in 2001? (4) And lastly, but perhaps most importantly, what have

---

1 Erling Lars Dale, *De strategiske pedagoger—Pedagogikkens vitenskapshistorie i Norge* (Ad Notam Gyldendal, 1999); Kim Helsvig, *Pedagogikkens grenser. Kampen om norsk pedagogikk ved Pedagogisk forskningsinstitutt 1938–1980* (Oslo: Abstrakt forlag, 2005), 324–339.

2 Kim Helsvig, ‘Kunnskapssamfunnet og nasjonen—den internasjonale vendingen’. *Nytt Norsk Tidsskrift*, 39.1 (2022), 5–17 <https://doi.org/10.18261/nnt.39.1>

been the changes and continuities in the use of psychometrics in the making of the Norwegian school from 1950 to 2025?

In the final part of the chapter, I therefore briefly discuss the question of whether the previously dominant, but for many years dormant, psychometric tradition in the Norwegian school and educational science has been revived in the post-2000 PISA era as a political tool for managing school and teacher accountability within a more general bureaucratization of educational policy. Alternatively, should the increase in the use of psychometrics and “IQ-like” tests since the beginning of the new millennium be understood as a tool to empower the pupils and adapt education to individual abilities and interests within the framework of the “knowledge society”?

## The Psychometric Tradition in Norwegian Educational Science

Norwegian educational science had a strong psychometric leaning from the very beginning, as exemplified by the career of Helga Eng, who, in 1938, took the chair as the first professor in educational science at the newly established *Pedagogisk forskningsinstitutt* [Pedagogical Research Department] at the University of Oslo. By then, she already had twelve years of experience as leader of the Psychotechnical Institute, from 1925 to 1937, where she had conducted intelligence and psychological aptitude tests on young students applying to industrial and vocational schools.<sup>3</sup>

From 1909, Eng had studied psychology and pedagogy for two years at the universities of Leipzig and Halle. During this time, she developed contacts with Ernst Meumann, William Stern, and Otto Lipmann. Meumann was the originator of the progressive pedagogical child-centered key concept “*vom Kinde aus*”, and was a leading figure in German experimental pedagogy. He served as professor and director at the first European institute for child and youth studies [*Jugendkunde*] in Hamburg from 1911 until his death in 1915. He was succeeded by his friend, William Stern, who from 1916 turned the institute into a leading European research center in psychology, pedagogy, and child and youth studies.

---

3 Elisabeth Lønnå, *Helga Eng—Psykolog og pedagog i barnets århundre* (Fagbokforlaget, 2002), 154–156.



Fig. 12.1 The psychologist and pedagogue Helga Eng (1875–1966) was a pioneer in developing psychometrics in Norway, first as leader of the Psychotechnical Institute from 1925, and, from 1938, as the first professor in pedagogy at the University of Oslo. Photographer unknown, Museum of University History, KHM, University of Oslo. All rights reserved.

Stern had organized intelligence research at his Institute for Applied Psychology in Berlin since 1905, together with his former student, Otto Lipmann. Stern also served on the board of the *Bund für Schulreform*, which was founded in 1908 and became an important source of inspiration for the establishment of the internationally very influential New Education Fellowship in 1920, championed by leading educationalists and philosophers of the time such as Carl Gustav Jung,

Jean Piaget, and John Dewey.<sup>4</sup> After her study period at this European center of progressive pedagogical research, Helga Eng wrote her dissertation, titled *Abstract Concepts in Children's Thought and Speech*. It was translated into German and published as a supplement in Stern and Lipmann's *Zeitschrift für angewandte Psychologie* [*Journal of Applied Psychology*], and Eng also benefited from her connections with Stern and Lipmann when establishing the Psychotechnical Institute in 1925.<sup>5</sup> She was also one of the main contributors to the Norwegian branch of the New Education Fellowship which was established in 1929.<sup>6</sup>

Just before becoming the pedagogy chair at the University of Oslo in 1938, Eng wrote, referencing the American educational philosopher John Dewey, that the task of educational research was to contribute to 'form[ing] people who can live in this changing and in many ways difficult world, people who, instead of being tossed around and destroyed, can guide and manage their life and behavior in such a way that it brings the greatest possible happiness to themselves and others'. According to Eng, pedagogy had to align itself with progressive societal development and recognize that people were living in a rapidly changing and increasingly technology-based reality. The departure from traditional and more static life contexts and societal structures imposed significant mental health challenges to the development of an integrated and secure personality, and the main task was therefore to help individuals become capable of actively engaging with a rapidly changing democratic society.<sup>7</sup> The Norwegian educationalist and historian of pedagogical ideas Erling Lars Dale highlights psychometric test research as one of Eng's three main areas for realizing this vision, the other two being research on thought and language, and research on personal growth based on biographical methods.<sup>8</sup>

---

4 Kevin, J. Brehony, 'A new education for a new era: the contribution of the conferences of the New Education Fellowship to the disciplinary field of education 1921–1938', *Paedagogica Historica*, 40(5–6) (2004), 733–755, <https://doi.org/10.1080/0030923042000293742>, 734.

5 Lønnå (2002), 83–85; Helsvig (2005), 42–43.

6 Ola Stafseng, 'Helga Eng og hennes pedagogikkfag 1938—mellom tradisjon og arvefølge', *Norsk Pedagogisk Tidsskrift*, 93.6 (2009), p. 438, <https://doi.org/10.18261/ISSN1504-2987-2009-06-03>. Helsvig (2005), 48.

7 Ibid.: 48, 66–73.

8 Dale (1999), 116.

Johannes Sandven—Eng's successor from 1948 to 1972—followed and developed Eng's psychometric work, making psychometrics the defining trait of the Pedagogical Research Department, and, in fact, of Norwegian pedagogical research in the first decades following World War Two. In 1947, Sandven spent six months at Teachers College, Columbia University, considered the nucleus of progressive pedagogy in the United States, to qualify for the professorship in pedagogy. Upon his return, he wrote *Pedagogisk idébrytning i USA* [*Educational Ideological Debates in the USA*]. Here, he paid special attention to the works of the Educational Policies Commission, which published four reports on education in the US between 1937 and 1946, and to a report presented by the Harvard Committee on educational policy in 1946.

The titles of these reports were: *The Unique Function of Education in American Democracy* (1937), *The Purposes of Education in American Democracy* (1938), *The Education of Free Men in American Democracy* (1941), *Policies for Education in American Democracy* (1946), and the Harvard Committee's *General Education in a Free Society* (1946). These titles make clear that the overarching context for pedagogy in the US during this period was the relationship between education and democracy. To support democratic development, these reports argued, schools had to ensure that pupils developed self-realization, human relations, efficiency, and civic responsibility. This was largely derived from a key passage in John Dewey's *Democracy and Education*. Here, Dewey emphasized that the school's most important tasks were to promote the pupil's ability to solve problems effectively, as well as their sociability, aesthetic taste, and conscientiousness.<sup>9</sup> In Sandven's language, these objectives were summarized in the phrase: 'Functional competence in relation to life's problems'.<sup>10</sup> In 1953, he reformulated these goals within a Norwegian

9 John Dewey, *Democracy and Education* (New York: The Macmillan Company, The Free Press, 1946), pp. 285–286. Here Dewey writes: 'We may say that the kind of experience to which the work of the schools should contribute is one marked by executive competency in the management of resources and obstacles encountered (efficiency); by sociability, or interest in the direct companionship of others; by aesthetic taste or capacity to appreciate artistic excellence in at least some of its classic forms; by trained intellectual method, or interest in some mode of scientific achievement; and by sensitiveness to the rights and claims of others—conscientiousness.'

10 Johs. Sandven, *Pedagogisk idébrytning i U.S.A.—En studie over utviklingen og de rådende synsmåter i dag* (Oslo: Fabritius og Sønners Forlag, 1949), p. 114.

context, writing that, 'to develop the child's *life competence* is one of our first and most important goals'.<sup>11</sup>

Sandven later articulated the contents of the goal to develop the life competence of children and young people through the concepts of: (1) efficiency (the ability for intellectual problem-solving, which was what IQ tests tried to measure); (2) consonance (developing emotional security that allows the child to express a rich and varied emotional life); and (3) character (developing a reason-guided, conscientious social attitude towards the community).

This tripartite division suggests that if we view life competence, the ability to live life well, as a unifying goal, it is important to shape the environment in which children grow up and the influences they are exposed to in such a way that it not only promotes proficiency in the intellectual domain, what is here termed efficiency, but also character, meaning the ability to integrate into society with others. At the same time, it is also essential to develop inner harmony and personal adjustment, what is here termed consonance, so that one achieves the greatest possible sense of security and freedom from inner conflict. [...] As society becomes more complex through technological development, and it becomes harder for the individuals to find their place within the whole of society, it becomes necessary for the educational process to reach as broadly and deeply into the personality as possible. It must encompass not only the intellect but also the character, the ability to empathize and understand others, to actively participate in the community, and likewise the ability to experience the world in a rich and multifaceted way, to live in peace with oneself, and to feel secure and confident when facing the challenges that life presents. [...] If we fail to view our educational task from this broad perspective, we may end up missing our goal.<sup>12</sup>

The concept of life competence became the guiding principle for Sandven, and the institute's development of psychometric tools for measuring intellectual, emotional, and social functioning followed on from such interests. In developing the tests, Sandven drew inspiration from similar work in other Nordic countries through his participation in the Expert Committee for Nordic Cooperation on Educational

11 Johs. Sandven, *Målsetningen i folkeskolen*, Special edition of *Skole og samfunn*, no.3 (1953), 2.

12 Johs Sandven, 'Fremtidens skole under pedagogisk-psykologisk synsvinkel', Special edition of *Pedagogisk forskning*, *Nordisk tidsskrift for pedagogikk*, 3 (1958), p. 127.

Research and Experiments. This work was particularly influenced by the committee's long-standing chairman, Torsten Husén from Stockholm University, who emerged as an international leader in the field of psychometric research in the postwar era.<sup>13</sup> Sandven also drew significant inspiration from the American educational psychologist and test researcher Lee Joseph Cronbach.<sup>14</sup> Apart from the adaptation of the Stanford–Binet intelligence test which Gard Paulsen writes about in this volume (chapter 9), Erling Lars Dale identifies four major and dominant psychometric research areas at the Pedagogical Research Department under Sandven from 1948 to around 1970:

1. The development of maturity tests for children aged six to fifteen, and the investigation of the predictive value of these tests in relation to school progress.
2. The development of psychometric instruments to assess pupils' emotional security and social attitudes.
3. The development of new evaluation methods, both in schools and in teacher education.
4. Social pedagogical research in schools, particularly focused on pupils' reactions to the school's working methods and on the factors influencing pupils' motivation and well-being in school.<sup>15</sup>

As we will soon see, the first of these areas—the maturity tests—would receive serious political attention and have an impact from the 1950s.

The maturity tests came to consist of six series: two for the age range from six to eight years, two for the nine to eleven age range, and two for

---

13 Interview with Lee Joseph Cronbach in Palo Alto, California, 28.02.2001. See also Torsten Husén, *En obotlig akademiker—En professors memoarer* (Stockholm, 1981) and Daniel Pettersson and Thomas S. Popkewitz, 'A chimera of quantifications and comparisons. The changing of educational "expertise"', in *New Practices of Comparison, Quantification and Expertise in Education*, ed. by Christina Elde Mølsted and Daniel Pettersson (London: Routledge, 2019), <https://www.doi.org/10.4324/9780429464904-2>

14 Interview with Per Rand, 13 November 2000. See also Per Rand, 'Hovedtrekk ved Johs. Sandvens forskningsinnsats', in *Pedagogikkens søkelys. Festskrift til professor Johs. Sandven på 70-årsdagen 1. november 1979* (Universitetsforlaget, 1979), pp. 1–18, 3.

15 Dale (1999), 192–216.



the ages from twelve to fifteen. They were meant to provide a picture of the intellectual functions and ability factors that were 'characteristic of the individual with special emphasis on those that are important in school'.<sup>16</sup> Each of the series for the six-to-eight-year age group was introduced with a test for hand movements and muscle coordination. The tests were constructed to measure short-term memory, long-term memory, perception of words, perception of sentences and situations, perception of room and form, perception of similarities, of opposites and of analogies, ability of judgment, and perception of quantities. As such, they covered the functional areas of memory, verbal ability, the ability to perceive room and form, the ability to reason, and the ability to perceive quantities.<sup>17</sup>

Sandven decided to call his new test apparatus 'maturity tests' and not IQ tests because of 'the many-sided and often misleading associations that have gradually come to be attached to the word intelligence'. The term 'maturity', on the other hand, had 'the advantage that it does not, like the term intelligence, lead our thoughts in the direction of something static, something that the person has or does not have in a certain quantity, something that can hardly be changed, but rather leads our thoughts towards something that is growing, something one has reached, but that is not necessarily the final stage'.<sup>18</sup> This was similar to the later, allegedly culturally neutral PISA tests that focused on the ability of fifteen-year-olds to use their reading, mathematical, and science knowledge and skills to meet real-life challenges, regardless of cultural setting and culturally and historically embedded knowledge.<sup>19</sup> Sandven emphasized that his maturity tests were 'constructed in such a way that they measure the fundamental maturity factors as far as possible, and are independent of knowledge and experience that are not in everybody's possession'.<sup>20</sup>

16 Johs. Sandven, 'Det teoretiske og metodiske grunnlag for modenhetsprøving: De norske modenhetsprøver for barne- og ungdomsalder', *Pedagogisk Forskning*, 6.1 (1962), 147–168, p. 165, <https://doi.org/10.1080/0031383620060110>

17 Ibid.: 166.

18 Ibid. (my italics).

19 PISA: Programme for International Student Assessment | OECD. <https://www.oecd.org/en/about/programmes/pisa.html>

20 Sandven (1962), 166.

## Psychometrics, Swedish Ideals and a New Norwegian School

From an international perspective, by the outbreak of World War Two, Norway had made significant moves towards an extensive, common elementary school system. There was no differentiation into various types of schools until after the completion of the compulsory seven-year primary school. After primary school, pupils could either choose to continue in a one- or two-year practically oriented continuation school [*Framhaldskolen*] or in the theoretically oriented middle school and high school [*Realskolen*]. In many other Western European countries, it was common to choose between different types of schools at an earlier age, which sent clear status signals and also strongly influenced further education.<sup>21</sup>

Nevertheless, there were significant differences within the common Norwegian elementary school, particularly between rural and urban areas. For example, due to low population density, classes in small rural schools were often “multigrade”, meaning children of different age levels attended the same class. The trend was clear—in urban municipalities, most pupils continued their education, often in status-giving theoretical studies. In rural municipalities, the children primarily attended the continuation school or went into unskilled work. This was seen as both economically and socially undesirable by new school policy actors in the Labor Party after the war. Led by Helge Sivertsen—State Secretary (deputy minister) from 1947, Director of Education in the capital Oslo for several decades, Minister of Education from 1960 to 1965, and the Labor Party’s undisputed leader in the school policy field from 1947 until around 1980—the new political school forces in the party would draw strong inspiration from Sweden.<sup>22</sup>

The new school policy line within the Norwegian Labor Party viewed the further expansion of the existing system of continuation schools and theoretical schools as irrational, considering overarching goals for the economically efficient utilization of the population’s knowledge resources and for national integration, both between social classes and between urban and rural areas. In their opinion, further development

21 Hans-Jørgen Dokka, *Reformarbeid i norsk skole*, NKS-Forlaget (1986), 12.

22 Helsvig (2005), 113–121.

of the old system would cement existing educational divisions and perpetuate societal inequalities. The main objectives of introducing a nine-year comprehensive school were therefore to improve conditions in rural municipalities, and to provide more pupils with the opportunity to pursue higher education by building larger schools that would incorporate both the continuation schools and theoretical schools. With this approach, the party also saw the potential to break down the existing status differences between practical and theoretical education by offering a division into general-practical and general-theoretical tracks during the last two school years under the same roof.<sup>23</sup>

By the end of 1940, a state educational policy committee had been established in Sweden with the ambition of creating greater coherence in the Swedish school system.<sup>24</sup> However, in 1944, the Labor Movement's postwar programme was introduced, in which the Swedish welfare state pioneer, and later United Nations diplomat and Nobel Peace Prize laureate Alva Myrdal advocated for the far more radical goal of a nine- or ten-year unified school that would eliminate all social, economic, and geographical barriers to education.<sup>25</sup> The Minister of Ecclesiastical Affairs and Education, Tage Erlander, soon established a new school commission, where those involved in drafting the proposal for the expanded and unified school were given key positions.<sup>26</sup> The core question was whether primary school pupils and middle school pupils could be taught together, not just in the same school, but in the same class. For this reason, the commission asked the National Institute of Psychological and Educational Research—which was established in 1944 under the leadership of John Elmgren, the professor of pedagogy at Gothenburg University—to conduct a study on 'the intelligence structure of Swedish youth'.<sup>27</sup>

Psychometric tests now came to play an important legitimizing role in the experiments with a nine-year unified school in Sweden. It was important for the conservative political opposition that these experiments be conducted in collaboration with scientific institutions.

23 Dokka (1986), 33.

24 Sixten Marklund, *Skolesverige 1950–1975. Del 1—1950 års reformbeslut* (Skolöverstyrelsen och Liber Utbildningsförlaget, 1980), p. 43.

25 Ibid., 33–36.

26 Ibid., 55.

27 Ibid., 74. Elmgren's study was published as *School and Psychology* in 1952.

As the educationalist Sixten Marklund wrote, the experimental activities became ‘the lifeline everyone chose to follow’. There was an almost ‘breathless faith’ for these activities to solve, or at least clarify, all the issues related to the reform.<sup>28</sup> The organization, the division into different tracks, the roles of teachers, and the design of vocational and high school preparatory education in the new school were to be subjected to scientific experimentation. The conservative Liberal People’s Party was reassured by the fact that the parliamentary school committee had emphasized that the reform would be guided by ‘the methods of modern science, particularly those of the natural sciences’.<sup>29</sup>

The Swedish experiences became a great source of inspiration for the new generation of educational policymakers in the Norwegian Labor Party, particularly Helge Sivertsen, who was now in the process of shaping a new Norwegian school for the future, modeled after what he called ‘the most modern country in the Nordic region’.<sup>30</sup> In the spring of 1947, the secretary of the 1946 school commission of Tage Erlander—now Sweden’s Prime Minister—spoke to *Sosialistisk Skolelag* [the Socialist School Society] in Oslo. The ambitions were enormous. The Swedish government’s reform was not just a reform of the schools; it was a profound social and political reform of the entire society. The new school was to be the foundation for an ‘education for a new world. [...] Why shouldn’t it be possible, let’s say at the dinner table in a farmer’s home, to discuss the structure of atoms? Why shouldn’t bricklayer apprentices discuss the social conditions for Shakespeare’s character creation during their lunch break? There are no limits here’.<sup>31</sup> Later that year, Helge Sivertsen gave a lecture on Norwegian educational policy at the same place, with much the same message.<sup>32</sup>

In 1949, Sivertsen then took the initiative to establish a dedicated group for psychology, education, and youth issues within the newly established Norwegian Research Council for Science and the Humanities (NAVF). It is reasonable to assume that he was inspired by the role that

---

28 Ibid., 224

29 Quote from Marklund (1980), 226.

30 Helge Sivertsen, ‘Norsk skolereform’, *Forsøk og reform* (Forsøksrådet for skoleverket, 1956), p. 15.

31 From Kim Helsvig, *Reform og Rutine. Kunnskapsdepartementets historie 1945–2017* (Pax Forlag, 2017), p. 59.

32 Tønnes Sirevåg, *Den ni-årige skolen* (Aschehoug, 1979), p. 13.

educational science was beginning to play in the Swedish reforms. In 1950, Helge Sivertsen was present as an observer during the Swedish parliamentary debate that led to the decision to experiment with a nine-year comprehensive school, and he pondered how something similar could be done in Norway before it was too late.<sup>33</sup>

## Psychometrics in Practice

In 1952, the opportunity arose for Sivertsen's ideas to be put into practice. In the final report from the government-appointed *Samordningsnemnda for skoleverket* [Coordination Committee for the School System], it was stated that further reform efforts should proceed on an experimental basis, and that this work should be extensive, systematic, and placed under expert leadership.<sup>34</sup> The Swedish inspirations were clear: 'There is certainly some merit in the path Sweden has chosen, [...] by thoroughly testing pedagogical experiments and innovations in the development of youth education through experimental schools and experimental programs'.<sup>35</sup>

The Church and Education Committee in the Norwegian Parliament then presented a report in 1954 that paved the way for close collaboration between educational research and new school policies:

The school system, by employing empirical science to test new school forms and pedagogical methods, will have brought the reform efforts in this important sector of society into modern forms. This method of testing is today used throughout the rest of society, and it should as well be a given within the school system.<sup>36</sup>

In this way, collaboration with pedagogical research provided legitimacy to decisions regarding extensive school policy reform experiments in both Sweden and Norway during the 1950s. The new school policy ideas gradually gained acceptance, largely due to the legitimacy that proponents of the new direction were able to give the reform proposals through an alliance with educational science.

---

33 Knut Jordheim, 'Okkupasjonsårene 1940–45 som "epoke" i norsk skoles utvikling' in *Skolen 1999–2000—Årbok for norsk utdanningshistorie*, ed. by Knut Jordheim (Stiftelsen SKOLEN—Årbok for norsk utdanningshistorie, 2000), pp. 58–76, p. 74.

34 Dokka (1986), 23.

35 Quote from Sirevåg (1979), 23.

36 *Innstilling O nr.60, 1954* in Sirevåg (1979), 25.

In 1950, Johannes Sandven's Pedagogical Research Department was given the first Research Council grant to develop the maturity tests for children aged six to fifteen, as described above. For the politicians, new tests were needed to differentiate pupils based on abilities and aptitudes as part of the trials of a nine-year, track-divided lower secondary school. According to Sandven, existing Norwegian tests were unable to measure the intellectual functions relevant to school. The tests were either culturally biased, outdated, or unsuitable for this purpose; additionally, the work that had begun under Helga Eng, and continued by Sandven's students, to develop a new Norwegian individual test based on the Terman–Merrill Stanford revision of 1937 had not yet been completed.<sup>37</sup>

However, there are indications that key educational policymakers and officials within the school administration were skeptical about the extent to which Sandven and the Pedagogical Research Department would be able to deliver what was needed. As a result, there were considerations of establishing a “test service center”, as in Sweden. Although it was suggested that the Norwegian Defense Research Establishment had the necessary expertise, and that the most desirable solution would be to establish a test department directly under the ministry, there was also a concern that it would be too much of an insult not to consult the Pedagogical Research Department, allowing it to ‘demonstrate what it is capable of achieving’.<sup>38</sup>

At the end of 1954, the head of *Forsøksrådet for skoleverket* [the School Reform Council], which Helge Sivertsen had established a few months earlier to supervise the reform, wrote a letter to Johannes Sandven inviting him to meet with the Minister of Church and Education. The purpose was to discuss the use of various selection methods for differentiation in the new school system.<sup>39</sup> The request from the politicians was clear: ‘What we primarily need tools to measure are: a) intelligence level, b) school readiness, c) aptitude for English, d) knowledge level in arithmetic and Norwegian after seven years of schooling, e) pupil profile related to selection for the various tracks in secondary school’.<sup>40</sup> The school authorities desired ‘a fixed agreement with

37 Sandven (1962): 148. See also Gard Paulsen, chapter 9, this volume.

38 From Helsvig 2005, 124.

39 Letter from Tønnes Sirevåg to Johs. Sandven, 20 December 1954. Kirke- og Undervisningsdepartementet. Forsøksrådet for skoleverket, 1251.2/10. Pedagogisk forskningsinstitutt, 1A.018.27, Box 8:3–2.

40 Letter from Tønnes Sirevåg to Ragnar Rommetveit, 8 October 1955 (Ibid.).

the institute for this practical research work so that the School Reform Council could request the institute to take on various tasks as the need arises'.<sup>41</sup>

In his reply, Sandven reminded the School Reform Council leader that 'motives and interests are environmentally labile conditions that undergo significant changes during adolescence'.<sup>42</sup> In this way, the researcher poured some cold water on the bureaucrat's scientific optimism, but the Pedagogical Research Department nonetheless became a kind of test service center for the Ministry of Church and Education for a few years from the latter half of the 1950s. The institute's psychometric maturity tests—which in practice were school-adapted IQ tests, although Sandven explicitly focused on the dynamic nature of the "maturity" that was measured—became a crucial legitimizing factor for developing a unified and track-divided nine-year school, modeled after the widely acknowledged educational and social-democratic pioneer country of the day, Sweden.

The school maturity tests were highly useful for the Labor Party during a critical reform phase from the mid-1950s, as they provided a plausible basis for differentiating between pupils within a tracked and expanded elementary school system. Until 1959, a total of nineteen elementary schools were part of a pilot reform programme.<sup>43</sup> Karl Mønnesland's IQ test from 1936, Sandven's maturity tests, and his colleague Ellen Nymoen's adapted Swedish 'Test for Theoretical Study' were used in connection with the differentiation of pupils after seventh grade in these initial experimental schools.<sup>44</sup>

## Critique, Decline, and Fall

However, while the other political parties, including the conservative opposition, in around 1960 almost unanimously supported the organizational reforms towards a non-tracked, unified nine-year comprehensive school, the tests had outlived their political use. At this point, public debate began to question the ethical basis of invasive educational research and the desirability of using educational tests in schools, as Fredrik Thue illustrates (chapter 11, this volume) about the

41 Letter from Tønnes Sirevåg to Johs. Sandven, 15 October 1955 (Ibid.).

42 Letter from Johs. Sandven to Tønnes Sirevåg, 22 October 1955 (Ibid.).

43 Melding om forsøk i skolen 1959–60. *Norsk skole. Opplysnings- og kunngjøringsblad for skoleverket*, nr.6, 1962, 106.

44 Ibid., 112–113.

1959 controversy over IQ testing of Norwegian schoolchildren. By then, pedagogues, teachers, philosophers, historians, and religious activists affiliated with the strongly Christian conservative *Institutt for Kristen oppseding* [Institute for Christian Education (ICE)] had already been criticizing the research at the Pedagogical Research Department for several years. The Inner Mission had a strong institutional foothold at Oslo Teacher's College, and as the ICE's first director, Bjarne Hareide, stated in 1950, the Christian conservatives strongly criticized the psychological and educational sciences 'which know so well how everything should be', and which aim to 'help modern humans achieve a scientific basis for their egocentricity' so that 'the future human being can live enclosed within themselves'.<sup>45</sup>

Outside the circle of University of Oslo educationalists, school professionals with different approaches to the field as well as much of the broader public were highly critical of the tests. The conservative opposition in Parliament drew widespread support from the public and from the educational community surrounding ICE for the view that the Pedagogical Research Department's research posed a threat, both to the integrity of families and individuals, as well as to Christian core values. Henceforth, the development of increasingly sophisticated and invasive tests of personal and social factors affecting the pupils' academic performance and socialization became highly controversial.

For Sandven, differentiation and individualized teaching were overarching educational goals. But he also urged caution in the use of the tests. He emphasized, in line with his concept of "life competence", that the 'intellectual, social, and emotional functions interact and form a whole where there is internal interplay and dependency', and that variations in personality pattern should be taken into consideration, 'so that no child is unduly forced'.<sup>46</sup> Therefore, one had to be careful when advising on school start and track selection, and it was the parents and the student who had to make the final decision:

The conclusion that can be drawn is that the test subject, with a certain degree of probability, will or will not be able to achieve the progress in question, which may be considered the goal from the parents' or the pupil's perspective. In my view, it is very important to make it clear to

45 Bjarne Hareide, 'Focus', *Prismet* 8/1950, IKO-forlaget (1950), p. 237.

46 Sandven (1962), 167.



those being guided that a categorical statement is not possible and that it is the degree of probability that can be presented. [...] This will, in many cases, be of great help to parents and students when it comes to reaching a sensible decision.<sup>47</sup>

For the Labor Party, the tests were primarily a means to legitimize structural reforms of the school system, and as these reforms had gained widespread support by 1960, the tests became far less politically relevant. However, the school maturity tests continued to be used in Norwegian schools well into the 1970s—but from the 1960s, they were primarily used as tools to guide the parents' and pupils' decision about whether a child was mature enough to start school at the age of seven, and not as tests for sorting or advising pupils for the final years of the nine-year comprehensive school. The latter became far less important in 1961, when the Parliament warned against differentiation in the new school being too early, and in 1963, it abandoned the formerly salient goal of a tracked lower-secondary school system.<sup>48</sup>

By the beginning of the 1960s, the tests had thus lost much of their political significance. In the following years, both the emerging economics of education within the Organization for Economic Co-operation and Development (OECD) and the Norwegian version of the international positivism critique played significant roles in further delegitimizing the use of psychometric tests in Norwegian schools. Educational economist Kjell Eide became Director General of the new Planning Department in the Ministry of Church and Education in 1964. From the outset, he was a strong opponent of the test research at the Pedagogical Research Department. During his time as a consultant for the OECD from 1961 to 1964, Eide led the secretariat for the 'Study Group in the Economics of Education', a think tank that brought together researchers and education planners. It became, in the words of George Papadoupoulos—OECD historian and longtime deputy director in the OECD's Directorate for Social Affairs, Management and Education—'a focal point and a prime mover in the development of a new and exciting branch of economics, the economics of education'.<sup>49</sup>

---

47 Ibid., 163.

48 Ibid., 163; White paper to the Parliament, *Nr. 22 (2010–2011), Motivasjon—Mestring—Muligheter* (Ministry of Education and Research, 2011), Meld. St. 22 (2010–2011) - regjeringen.no, 112–113. See also Fredrik Thue, chapter 11, this volume.

49 George S. Papadoupoulos, *Education 1960–1990—The OECD perspective* (OECD,

Throughout the 1950s, Eide had worked together with the Minister of Church and Education to create a closer cooperation between educational research and the educational policy administration within the ministry. However, this work had led to little more than a growing sense of disappointment among the ministry's leadership that the psychometric educational research had 'little relevance to the practical day-to-day life in schools'.<sup>50</sup> This dissatisfaction was largely directed specifically at Sandven and the Pedagogical Research Department. As Eide later stated:

By the mid-1960s, educational research in Norway had reached a certain scale, primarily concentrated at the Pedagogical Research Department at the University of Oslo. Professor Johs Sandven had done pioneering work in establishing the institute. However, under his dominant leadership, the institute adopted a research tradition that offered a rather narrow definition of the issues, and neither schools nor the ministry found the research particularly relevant to practical problems.<sup>51</sup>

As Director General, Kjell Eide steadily advocated for more practically and politically oriented educational research. A former OECD bureaucrat, Eide was particularly inspired by the US and Sweden, where there was now a strong focus on 'more "product-oriented" research and development work'.<sup>52</sup> In his critique of the test research, Eide used a broad definition of the concept of educational research, which also comprised 'activities that could be characterized as development work or practical experiments'. Therefore, in addition to pedagogy and psychology, educational research in Norway also had to include contributions from sociologists and economists, and to some extent, social anthropologists and political scientists.<sup>53</sup>

The Director General's advice for the future was clear and in line with recommendations from the OECD—there was a need to increase political and administrative control over research, and it was necessary to build up the Ministry's own educational research management competence.<sup>54</sup>

---

1994), pp. 41, 56, footnote 6.

50 Letter from Kjell Eide to Alfred Oftedal Telhaug. From Telhaug, Alfred Oftedal, *Forsøksrådet for skoleverket (1954–1984)* (Universitetsforlaget, 1990), 169, footnote 60.

51 Kjell Eide, *Departementets lille kanarifugl, eller kulturpolitikk bli til* (NIFU, 1985), 199.

52 *Ibid.*, 18.

53 *Ibid.*, 10–12.

54 *Ibid.*, 22.

This led to the creation of a special chapter for educational research in the state budget in 1970, and from 1973, this chapter was placed directly under the ministry and Eide's control in the Planning Department. In this way, the ministry effectively cut off much of the financial support for Sandven and the test research.<sup>55</sup>

Sandven had by then presented a cautious but fundamental critique of the development of the Norwegian school following the Labor Party reforms. From 1966 to 1969, he led a longitudinal study of 3,500 students based on the totality of the psychometric tools that were developed at the Pedagogical Research Department. It appeared that the Norwegian school primarily provided good and predictable academic development for pupils with low emotional security (consonance), a strong sense of social duty (character), and below-average problem solving capacity (efficiency):<sup>56</sup> 'A school system that in principle largely has the same structure and work patterns from 1<sup>st</sup> to 12<sup>th</sup> grade can hardly be said to be sufficiently adapted to the developmental changes and differences among pupils'.<sup>57</sup> As such, Sandven's psychometric programme seemed to be at odds with the development of the Norwegian school more broadly.

It was, however, the *Positivismusstreit* that dealt the final blow to the primary- and lower-secondary-school-related test research that had been developed in Norway up until around 1970. Towards the end of the 1960s, the government's proposal for fundamental higher education reform faced strong opposition from a radicalized and growing student body that perceived the proposal as an attempt to make higher education more efficient for the interests of state and industry in a way that undermined opportunities for reflection and intellectual growth.<sup>58</sup> This opposition was largely fueled by the rising critique of positivism within the humanities and social sciences, which for some time had been championed in Norway by the philosopher Hans Skjervheim.<sup>59</sup>

55 Helsvig 2005, 205–211.

56 Dale (1999), 209–216.

57 Johs Sandven, 'Årsaker til manglende skoletrivsel', *Forskning med relation til utbildning för åldersklasserna 16–19 år* (Nordisk Komité for Pedagogisk Forskning, 1971). The study followed the pupils through the three years of upper secondary school, thus the reference to 1<sup>st</sup> to 12<sup>th</sup> grade.

58 Fredrik Thue and Kim Helsvig, *Universitetet i Oslo. Den store transformasjonen* (Oslo: Unipub, 2011), pp. 328–338.

59 Ibid., 334.

Since delivering his thesis for the Magister degree *Objectivism and The Study of Man* in 1957, Skjervheim had become the leading advocate in Norway for an international critique that the contemporary social sciences were overly influenced by a natural science mode of thinking. Skjervheim primarily criticized the belief that studies of human relations could also strive for an objective spectator role, arguing that such an approach was epistemologically invalid and could have politically disempowering and authoritarian implications. According to Skjervheim, the social scientist who was inspired by natural science research ideals overlooked the fundamental condition for understanding human relations—that it was a community of communication in which the researcher was also a participant. Skjervheim emphasized that a social scientist would make a fundamental epistemological error by creating an analytical divide between themselves and the research object, as in the natural sciences, and that, in practice, it was impossible for the social scientist to escape the hermeneutic circle.<sup>60</sup>

The core theme of this critique involved the conditions for understanding and learning about human relations, and for Skjervheim, ‘the fundamental problems in educational philosophy were the same as the fundamental problems in philosophy, plain and simple’.<sup>61</sup> He criticized Norwegian pedagogy in general for being too heavily influenced by natural scientific modes of thinking, and even though Skjervheim knew little of the actual research at the Pedagogical Research Department, and his critique can be seen as both misguided and bombastic, this hit Sandven’s psychometric research programme hard as it was so heavily inspired by the natural sciences.<sup>62</sup>

The seminar that Skjervheim organized in his hometown of Voss, in the heart of Norway’s countercultural western region, towards the end of the year of student revolts in 1968 became the major catalyst for the critique against the Pedagogical Research Department. The seminar’s theme was Pedagogy and Society, and it strongly focused on the role of education in Norwegian society. Pedagogy now became the central object of the Norwegian positivism critique, and when the presentations

60 Hans Skjervheim, ‘Den eigne refleksjon og fortolkinga av andre’, in *Deltakar og tilskodar* (Tanum-Norli, 1976a), pp. 287–299, 298.

61 Hans Skjervheim, Eit grunnproblem i pedagogisk filosofi, 226–245 in *Deltakar og tilskodar* (Tanum-Norli 1976b), p. 243.

62 Helsing (2005), 222–227 and 257–263.

from the seminar were published, the book cover asked whether the school was making pupils 'victims of a technocratic-materialistic society'.<sup>63</sup> Within a few years, a wave of opposition built up against all forms of measurement, testing, efficiency, and rationalization in the school and education sector.<sup>64</sup>

The strict, natural science-inspired demands for rigor of the Sandven tradition, its power-oriented stubbornness, and its dismissive attitude in the face of increasing demands for value-based relevance, utility, and significance thus contributed, from the early 1970s, to unite the otherwise very different Christian-conservative, political-bureaucratic, and positivism-critical approaches to education in a joint attack on the test research. This alliance brought together a coalition of cultural forces that had previously, in Norwegian history, shown itself to be capable of bringing about significant changes—namely a popular and anti-authoritarian movement in alliance with conservative Christianity and ambitious bureaucrats and politicians. Like the alliances that united the Norwegian left-wing movement in the latter half of the nineteenth century against the jurist-dominated *Embetsmansstaten* (the civil servant state) and the opposition to the EEC/EU in the 1970s and 1990s, the anti-test alliance in pedagogy would also fracture as soon as the unifying enemy, in this case Sandven, was overcome.

Sandven was forced to step down as department director in 1972, and when he retired in 1980, his professorship was filled by one of his fiercest critics since the 1950s—a Christian conservative educationalist.<sup>65</sup> From the mid-1970s, therefore, the Pedagogical Research Department and Norwegian educational pedagogy became highly politicized and polarized between Christian-conservative educators who focused on knowledge, culture, and tradition, and anti-authoritarian educators who focused on well-being, participation, and socialization. However, both groups could agree that psychometric testing was a positivist misstep that belonged to the past in Norwegian schools. And until the turn of the century, they could count on the support of this view from both the civil servants in the ministry of Church and Education, and most Norwegian politicians.

63 Front cover of Hans Skjervheim and Leif Tufte, eds., *Pedagogikk og samfunn. Foredrag på Vosse-seminaret 13.-15-desember 1968* (Gyldendal, 1969).

64 Helsvig (2005), chapter 8. See also Jon Kyllingstad, chapter 10, this volume.

65 Universitet i Oslo. Det samfunnsvitenskapelige fakultet. 1251.3/10, bd.3, "0245, Stillinger, Professorat i pedagogikk (etter Sandven)—Torstein Harbo, 1978–81".

## The “Knowledge Society” and PISA

The use of psychometric tests in the Norwegian school would nevertheless return on a grand scale after the publication of the first PISA study in 2001, in the context of the emerging “knowledge society”. The most immediate roots of the knowledge society lie in the revitalization of human capital theory in the 1980s.<sup>66</sup> In 1983, the Reagan administration’s education report, *A Nation at Risk: The Imperative for Educational Reform*, warned against ‘a rising tide of mediocrity’ in the American school system.<sup>67</sup> *A Nation at Risk* set the tone for what would become the central theme in the international discourse on the knowledge society—that the school and education system must be more aligned with the needs of business to ensure the nation’s economic strength and power in the knowledge economy through more schooling, a focus on basic skills (‘The New Basics’), and standardized tests and assessments.<sup>68</sup>

Under the strong influence of Ronald Reagan’s US and Margaret Thatcher’s United Kingdom, the OECD in the 1980s turned to focus on neoliberal perspectives such as school choice in a free market, new funding sources, business connections, accountability, efficient resource use, and performance-based rewards for teachers.<sup>69</sup> The OECD has never had formal power over member countries, but its “soft power” has been significant.<sup>70</sup> The OECD collects comparative data, evaluates national school and education systems, and both advises on and monitors policy development.<sup>71</sup> The organization’s ambition is to help member countries ‘understand the relative strengths and weaknesses of their own

66 Helsvig (2017), 175.

67 Jal Mehta, ‘Escaping the Shadow: A Nation at Risk and Its Far-Reaching Influence’, *American Educator*, 39.2 (2015), 20–26, p. 20.

68 Ibid., 20–21.

69 Kjell Eide, *OECD og norsk utdanningspolitikk. En studie av internasjonalt samspill* (NIFU, 1995), p. 95.

70 Tore Bernt Sørensen, Christian Ydesen and Susan Lee Robertson, ‘Re-reading the OECD and education: the emergence of a global governing complex—an introduction, 101’ *Globalisation, Societies and Education*, 19.2 (2021), <https://doi.org/10.1080/14767724.2021.1897946>

71 Gustav E. Karlsen, ‘Internasjonale aktører som premissgivere for norsk utdanningspolitikk med særlig vekt på OECD’, in *Reformideer i norsk skole: Spredning, oversettelse og implementering*, ed. by K. A. Røvik, T.V. Eilertsen and E.M. Furu (Cappelen Damm akademisk, 2014); Børhaug, Kjetil and Kim Helsvig, ‘Kampen om samfunnsfaget—eit forskingsoversyn med vekt på lærarane si rolle’, *Nordidactica—Journal of Humanities and Social Science Education* (2023), p 37.

education system' through international comparative studies.<sup>72</sup> Since the 1990s, the OECD's influence on school and education policy, both within and outside OECD countries, has become increasingly strong.<sup>73</sup>

In 1996, the year before the PISA launch, the OECD wrote that knowledge had become the driving force behind productivity and economic growth.<sup>74</sup> According to the organization, all member countries were moving towards a knowledge-based economy and were therefore more concerned than ever with developing their human capital to boost economic growth, increase employment, and promote social integration.<sup>75</sup> In the OECD view, the success of nations, communities, and individuals depended on adapting to these changes.<sup>76</sup> Over time, the OECD developed educational economic models that posited a connection between pupils' PISA scores in reading, math, and science on one side, and economic growth on the other. According to these models, even a small PISA improvement of twenty-five points over a period of twenty years would yield an economic gain of an astronomical 115 trillion US dollars across the OECD area over the lifetime of a generation.<sup>77</sup>

The discourse on the knowledge society emerged from the international conservative wave of the 1980s, with Thatcher and Reagan at the forefront, but it also became a guiding principle for center-left politicians on both sides of the Atlantic in their subsequent search for a "third way" between socialism and capitalism after the fall of the Berlin Wall and the dissolution of the Soviet Union. In 1994, the Democrat US president Bill Clinton stated that 'what you earn depends on what you learn', and Tony Blair and New Labour's British path to economic

72 OECD, *PISA 2015 Results (Vol. II): Policies and Practices for Successful Schools* (OECD Publishing, 2016), p. 10.

73 Christian Ydesen and Karen Egedal Andreassen, 'Historical Roots of the Global Testing Culture in Education', *Nordic Studies in Education*, 40.2 (2020), p. 161. <https://doi.org/10.23865/nse.v40.2229>

74 OECD, *The Knowledge-Based Economy* (OCDE/GD, 1996), p. 3.

75 OECD, *Human Capital Investment: An International Comparison*, (OECD Publishing, 1998), 3. See also Helen Seitzer, D. Niemann and K. Martens, 'Placing PISA in perspective: the OECD's multi-centric view on education', *Globalisation, Societies and Education*, 19.2 (2021), 198–212, <https://doi.org/10.1080/14767724.2021.1878017>

76 OECD, *The Well-being of Nations: The Role of Human and Social Capital*, (OECD Publishing, 2001).

77 OECD, *The High Cost of Low Educational Performance: The Long-run Economic Impact of Improving PISA Outcomes* (PISA, OECD Publishing, 2010), p. 6.

growth and a better future for all in the second half of the 1990s focused on ‘education, education, education’.<sup>78</sup> Education was also at the core of the vision that Blair and German social democratic chancellor Gerhard Schröder outlined in 1998 for an innovative and forward-looking social democratic Europe: ‘The most important task of modernization is to invest in human capital: to equip both individuals and businesses for the knowledge-based economy of the future’.<sup>79</sup> By 2009, the ‘twenty-first-century skills’ promoted by PISA had been incorporated into the curricula of most OECD countries.<sup>80</sup>

Also in Norway, warnings about declining standards in schools initially came from the political right at the beginning of the 1980s.<sup>81</sup> However, it was the sociologist and Labor Party member Gudmund Hernes who truly brought the theme of *A Nation at Risk* to the forefront in Norway. In his 1986 op-ed ‘Is it possible to have ambitions in Norway?’, Hernes—writing from Harvard University—painted a picture of a mediocre University of Oslo at the top of a complacent and unambitious Norwegian school and education system.<sup>82</sup> Hernes became the Minister of Education, Research, and Church Affairs in Gro Harlem Brundtland’s third Labor government in 1991, and throughout the 1990s, the focus on education as the means to promote economic growth and international competitiveness became a cross-party goal in Norway.<sup>83</sup> As Hernes himself put it: ‘The future of Norway will not be driven forward by oil. The future of Norway will run on knowledge—a high-octane, unleaded, environmentally-friendly fuel, full of creativity, intelligence, and reliable facts’.<sup>84</sup> This message resonated across the

78 McKnight, A., H. Glennerster, and R. Lupton, ‘Education, education, education ... an assessment of Labour’s success in tackling education inequalities’ in *A more equal society?: New Labour, Poverty, Inequality and Exclusion*, ed. by J. Hills and K. Stewart (Bristol University Press, 2005), pp. 47–68.

79 Tony Blair and Gerhard Schröder, *Europe: The Third Way/Die Neue Mitte*. Working Documents (Friedrich Ebert Foundation, South Africa Office, 1998), p. 3.

80 Erlend Eidsvik, ‘Fagfornyng, berekraft og geografisk danning’, *Norsk pedagogisk tidsskrift*, 104.3 (2020), 268–282, p. 271; A. Nordin, and D. Sundberg, eds., *Transnational Policy Flows in European Education: The Making and Governing of Knowledge in the Education Policy Field* (Oxford: Symposium Books, 2014).

81 Helsevig (2017), 149–150.

82 Ibid., 179.

83 Alfred Oftedal Telhaug, A.I. Wallum and R.T. Tønnesen, *Norsk utdanningspolitisk retorikk, 1945–2000* (Cappelen Akademisk, 1999), 148–149.

84 In Ole Kolsrud, *Rekonstruksjon og reform. Regjeringskontorene 1945–2005* (Universitetsforlaget, 2008), p. 259.



Norwegian educational policy landscape, especially after the OECD's economic report on Norway in 1997: Norwegian school, education, and research was very expensive and delivered mediocre results at a time when human capital was deemed essential for economic growth.<sup>85</sup>

In 2000, the first Jens Stoltenberg government (Labor), echoing Blair and Schröder, called for action to meet a 'global knowledge economy' where 'knowledge and ideas will be our most important capital'.<sup>86</sup> In its inaugural declaration in the fall of 2001, a new center-right government stated that Norway should 'become a nation at the international forefront in terms of new technology, competence, and knowledge'. Their path, too, was paved with human capital: 'The values and attitudes of the population, their knowledge and skills, are [...] the greatest resource Norway has in facing the future'.<sup>87</sup>

The OECD's message was only strengthened by the "PISA shock" just over a month later when Norwegian pupils scored average results in the organization's new international school survey.<sup>88</sup> The now largely cross-party economic thinking in Norwegian school policy was crystal clear in the government school commission report from 2003, leading up to the major curriculum *Kunnskapsløftet* [Knowledge Promotion reform] in 2006. In the chapter 'Human Capital—Our Most Important Resource', it was stated that '[t]he competence of the population is now considered the most important single factor in a country's economic performance'.<sup>89</sup>

## The Revival of Psychometrics After PISA

In Norway, as in many other Western countries, the first PISA results were followed by an almost unanimous national consensus that the school system was in crisis. With a high degree of cross-party

85 OECD, *OECD Economic Surveys: Norway 1997* (OECD Publishing, 1997), chapter IV, pp. 119–123.

86 Norwegian government (2000), Regjeringen Stoltenbergs tiltredelseserklæring - regjeringen.no [https://www.regjeringen.no/no/aktuelt/regjeringen\\_stoltenbergs\\_tiltredelseserk/id263688/](https://www.regjeringen.no/no/aktuelt/regjeringen_stoltenbergs_tiltredelseserk/id263688/)

87 Norwegian government (2001), Regjeringen Kjell Magne Bondeviks tiltredelseserklæring - regjeringen.no [https://www.regjeringen.no/no/aktuelt/regjeringen\\_kjell\\_magne\\_bondeviks/id262243/](https://www.regjeringen.no/no/aktuelt/regjeringen_kjell_magne_bondeviks/id262243/)

88 Svein Sjøberg, 'PISA-syndromet—Hvordan norsk skolepolitikk blir styrt av OECD', *Nytt Norsk Tidsskrift*, 31.1 (2014).

89 NOU 2003: 16. *I første rekke—Forsterket koallitet i en grunnopplæring for alle*, NOU 2003: 16 - regjeringen.no, 33.

agreement, within a few years several measures were implemented to raise the quality of education and improve the PISA scores.<sup>90</sup> The most important of these was the establishment of The Norwegian Directorate for Education and Training in 2004.

For the first time in the 180-year long history of the ministry, responsibility for key areas of the central school administration was transferred from the ministry to a directorate. This move was intended to ease the ministry's workload and better enable it to function as a policy-formulating secretariat for the political leadership. The directorate was tasked by the ministry—which, in line with the international ideology of the knowledge society in 2006, was renamed, literally in Norwegian, the Ministry of Knowledge<sup>91</sup>—with responsibilities including curriculum development, quality assessment, school development, competency development, research, and statistics. In short, the primary task of the Norwegian Directorate for Education and Training was to facilitate the improvement of the Norwegian PISA results. The directorate now became 'the absolute authoritative center' in the Norwegian education sector, as the political scientist Kjell Arne Røvik and his colleagues wrote.<sup>92</sup>

From the outset, the Directorate for Education and Training started the development of a National Quality Assessment System (NQAS). From its inception in 2004, the system has gradually expanded to include thirty-five indicators, such as national tests, *kartleggingsprøver* [diagnostic tests], *skolebidragsindikator* [school contribution indicators], pupil surveys, as well as international surveys such as PISA, TIMSS (mathematics and science), PIRLS (reading), and ICCS (civic understanding).<sup>93</sup> All of these are psychometrically based measurements and surveys, and they form the backbone of the NQAS.<sup>94</sup>

90 Helsvig (2017), 230–236.

91 Formally in English, Ministry of Education and Research.

92 Kjell Arne Røvik, Tor Vidar Eilertsen and Torbjørn Lund, 'Hvor har de det fra, og hva gjør de med det? Utdanningsdirektoratet som innhenter, oversetter og iverksetter av reformideer', in *Reformideer i norsk skole Spredning, oversettelse og implementering*, ed. by Kjell Arne Røvik, Tor Vidar Eilertsen and Eli Moksnes Furu (Oslo: Cappelen Damm, 2014), pp. 88–118, 88.

93 NOU 2023: 1. *Kvalitetsvurdering og kvalitetsutvikling i skolen. Et kunnskapsgrunnlag*, NOU 2023: 1 - regjeringen.no, 76.

94 Sigrid Blömeke, and Rolf Vegar Olsen, 'På vei mot et sammenhengende nasjonalt kvalitetsvurderingssystem', *Acta Didactica Norge*, 12.4 (2018), 5–9 <https://doi.org/10.5617/adno.6278>

The catalyst for the establishment of the NQAS was the PISA survey. As we have seen, this survey was driven by much the same motivation as Johannes Sandven's maturity tests from the 1950s. They both aimed to measure fundamental school-relevant skills that were allegedly culturally neutral and, as Sandven put in the 1950s, 'independent of knowledge and experience that are not in everybody's possession'.<sup>95</sup> Sandven's tests were, for all practical purposes, school-oriented IQ tests, although he chose to call them maturity tests to emphasize the dynamic element of what was being measured. PISA, on the other hand, was formally designed to 'measure students' capacity to apply knowledge and skills in key subject areas and to analyse, reason and communicate effectively as they pose, solve and interpret problems in a variety of situations that can take place in a mathematics, reading, or science context'.<sup>96</sup>

However, a 2022 study of the 2018 PISA test in thirty-three OECD countries, conducted by a group of researchers in education, sociology, philosophy, cognitive science, and assessment, points out that this PISA definition is almost identical to prominent definitions of intelligence, i.e., the 'ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought'.<sup>97</sup> The research group notes that 'both achievement tests and intelligence tests include similar subsets of items, for example assessing vocabulary, reading comprehension, mathematical concepts, numerical skills'.<sup>98</sup> With reference to David Armor's 2003 anthology *Maximizing Intelligence*, they suggest that the substantial overlap between IQ and achievement tests, although very different in design, indicates that they are measuring more or less the same thing: 'general reasoning skills'.<sup>99</sup> The study further refers to fifty-four articles that have investigated the relationship between IQ tests and PISA, finding an overall correlation of 0.88. This

95 Sandven (1962), p 166.

96 OECD, *Science Competencies for Tomorrow's World. Volume 1* (Paris: Organisation for Economic Co-operation and Development, 2007), 16; Arthur Pokropek, Gary N. Marks, Francesca Borgonovi, Piotr Koc and Samuel Greiff, 'General or specific abilities? Evidence from 33 countries participating in the PISA assessments', *Intelligence*, Volume 92 (2022), <https://doi.org/10.1016/j.intell.2022.101653>

97 Pokropek et al. (2022).

98 Ibid.

99 Ibid.

means that, on average, about eighty percent of the variation in the PISA studies can be attributed to general cognitive ability.<sup>100</sup>

Since these cognitive abilities are school-related problem-solving skills, it is no surprise that PISA primarily produces the rather trivial knowledge that, for all practical purposes, the Western countries score equally, that developing countries score much lower than all the rest, and that Japan, China, and South Korea beat everyone. As the mathematician and economist Jan Ubøe says, this might have some obvious explanations that we do not need PISA to understand. The Western countries have had comparable cultural trajectories and run their school systems according to predominantly the same principles, and they all struggle with school-weary and unmotivated students and have a mixed group of good and less good teachers. Developing countries have poorly developed school systems and have much to gain from putting these in place, whereas PISA-winners Japan, China, and South Korea have effectively organized society so that education is the most promising way out of poverty, making students and parents throw everything else overboard to work hard for good school and test results, often through supplementary education.<sup>101</sup>

This suggests that both Sandven's maturity tests from the 1950s and PISA largely measure the pupils' general school-related cognitive ability. For Sandven, this was the explicit goal, whereas PISA claims to measure pupils' knowledge and skills across national education systems as a basis for an international ranking of educational performance with enormous influence and significance in educational policy, not just within the OECD area. PISA is supposed to measure school results across national borders, but to do so, they must be able to abstract "something" out of the school subjects that is not linked to national curricula or scholarly traditions and cultures. To achieve this, the PISA survey translates disciplinary knowledge in subjects such as mathematics, science, and history into "literacies" that have little to do with the disciplines themselves, but which are 'discrete concepts, skills, and information activated and articulated as objects of pedagogy, psychologies of

---

<sup>100</sup> Ibid.

<sup>101</sup> Jan Ubøe, 'PISA har utspilt sin rolle', *NHH-Bulletin* (2016), PISA har utspilt sin rolle | NHH. <https://www.nhh.no/nhh-bulletin/artikkelarkiv/eldre-saker/2016/desember/pisa-har-utspilt-sin-rolle/>

learning, and psychometrics’—what the curriculum historian T. S. Popkewitz calls the alchemy, the ‘translations of disciplinary knowledge into student literacies’.<sup>102</sup>

The educationalist William C. Smith notes that the ‘reinforcing nature of the global testing culture leads to an environment where testing becomes synonymous with accountability, which becomes synonymous with education quality’.<sup>103</sup> This global testing culture is also closely connected to what Finnish educationalist Pasi Sahlberg calls the Global Education Reform Movement (GERM), which ‘broadly follows the tenets of New Public Management and Neoliberalism’.<sup>104</sup> In this way, PISA has become, as PISA director Andreas Schleicher has said, ‘essentially a story of how international comparisons have globalized the field of schooling and education, which we have usually considered a domain of national policy’.<sup>105</sup>

PISA and the testing culture nevertheless possess a history that goes back to at least before World War Two, to the Bund für Schulreform, the New Education Fellowship and the scholarly efforts and environments—to which the Pedagogical Research Department’s first professor, Helga Eng, also belonged.<sup>106</sup> It is reasonable to draw a historical line between the American psychologist Edward L. Thorndike (1874–1949), the Swedish psychometrician Torsten Husén (1916–2009)—Johannes Sandven’s contemporary inspiration—and the German-born PISA director Andreas Schleicher (1964) in creating an educational thinking ‘based on a chimera of quantifications and comparisons’.<sup>107</sup>

## Management or Empowerment?

As in other Scandinavian and Western European countries, after PISA and over the past twenty years, Norway has developed management-oriented tests and curricula. Norway is thus part of a large international

102 Thomas S. Popkewitz, ‘Infrastructures and phantasmagrams of inclusions that exclude: international student assessments’, *International Journal of Inclusive Education*, 29.8 (2023), 1424–1444, <https://doi.org/10.1080/13603116.2023.2275156>

103 Ydesen & Egedal Andreasen (2020), 150.

104 Ibid. See also Seitzer et al. (2021), 202.

105 PISA leader Andreas Schleicher in TEDTalk 21.2.2013: Andreas Schleicher: Use data to build better schools - YouTube. <https://www.youtube.com/watch?v=7Xmr87nsl74>

106 Ydesen & Egedal Andreasen (2020), 150–154.

107 Pettersson & Popkewitz (2019), 19.

trend that emphasizes documenting educational processes and managing them, according to The Norwegian Directorate for Education and Training.<sup>108</sup> In an international comparison, the Norwegian bureaucratic school management system is furthermore one of the most centralized.<sup>109</sup>

During the preparations for NQAS, it was envisioned that the quality assessment system would be both a pedagogical tool for teachers working with their pupils and a governance instrument for authorities. As late as 1999, the parliamentary report 'Towards richer goals' [*Mot rikere mål*] stated that the overarching goal of the quality assessment work was to improve education for the pupils, and that it 'should not be used as a basis for external control and ranking of educational institutions'.<sup>110</sup>

However, after the first PISA survey in 2001, the focus clearly shifted toward making the quality assessment system primarily a tool for providing management information to school authorities. The overarching goal now was to improve PISA results and climb the rankings, as this was accompanied by promises of significant future economic gains for the nation. All the evaluations of NQAS point to the same conclusion—the quality assessment system has primarily provided the output results by which the Ministry and the Directorate manage the Norwegian school. The verdict is clear: the system the Directorate developed from 2003/2004 has primarily been a management instrument for the Ministry and the Directorate itself.<sup>111</sup>

At the same time, NQAS developed as it went along, perhaps also on somewhat feeble academic grounds. In the 1950s, Norwegian school authorities could engage with a pedagogical research community that had been working continuously for decades, and in close cooperation

108 Utdanningsdirektoratet, *Delrapport, Nasjonalt kvalitetsvurderingssystem* (2010), Nasjonalt kvalitetsvurderingssystem | udir.no

109 Peter Allerup, Velibor Kovac, Gro Kvåle, Gjert Langfeldt and Poul Skov, *Evaluering av det Nasjonale kvalitetsvurderingssystemet for grunnsopplæringen* (FOU-rapport 8/2009, Agderforskning and Aarhus: The Danish School of Education, Aarhus University for The Norwegian Directorate for Education and Training, 2009), p. 275.

110 White paper to the Parliament, *Nr. 28 (1998–99), Mot rikere mål* (1999), St.meld. nr. 28 (1998–99) - regjeringen.no, chapter 5, 7; Gjert Langfeldt and Torunn Lauvdal, *Evaluering av det nasjonale kvalitetsvurderingssystemet. Delrapport pr. 1.4.2006* (The Norwegian Directorate for Education and Training, 2006), *Evaluering av det nasjonale kvalitetsvurderingssystemet - delrapport 2006*, 14.

111 NOU 2023: 27, *Et nytt system for kvalitetsutvikling—for elevenes læring og trivsel*, 7; Allerup et al. (2009), 13–14; Langfeldt and Lauvdal (2006), 15–22.

with an international research community, to develop psychometric instruments as part of the efforts to create a new school system. This would be in the spirit of the Bund für Schulreform and the New Education Fellowship—more tailored to empower pupils and enhance their abilities according to their talents and interests. In stark contrast to this, as late as 2008 and four years into the NQAS, the Ministry of Education and Research called for an initiative to build up Norwegian research competence in educational measurement, claiming that there had never been a tradition of this kind of research in Norway.<sup>112</sup> This, firstly, showed that the Ministry of Education was completely ignorant of the dominant psychometric tradition of Norwegian educational science from the founding of the Pedagogical Research Department in 1938 until around 1970. Secondly, it also bore great historical irony—the Ministry was now trying to revive a research tradition in Norwegian educational science that it had nearly put to sleep about four decades earlier.

In 2004, the Department of Teacher Education and School Research at the Faculty of Educational Sciences at the University of Oslo was given the task to create the first national tests. The work immediately received criticism, both for primarily being focused on the pupils' final competence and not being supportive of learning, and for being almost identical to the PISA tests.<sup>113</sup> The evaluation of the national tests recommended a significant increase in competence in the field of test theory.<sup>114</sup> Only after very clear and active recommendations from the Ministry was the Centre for Educational Measurement (CEMO) founded at the University of Oslo in 2012, and over the past decade, the very internationally oriented CEMO has established itself as the leading psychometric research community in Norway. But the Department of Teacher Education and School Research at the University of Oslo continued to conduct not only the national test and supervise the PISA tests, it also had the responsibility for OECD's Teaching and Learning

112 Kunnskapsdepartementets avdeling for analyse-, internasjonalt arbeid og kompetansepolitikk, AIK (Kunnskapsdepartementet, 2008), 6.

113 Svein Lie, Marion Caspersen and Julius K. Björnsson, *Nasjonale prøver på prøve. Rapport fra en utvalgsundersøkelse for å analysere og vurdere kvaliteten på oppgaver og resultater til nasjonale prøver våren 2004* (Department of Teacher Education and School Research, University of Oslo and Námsmatsstofnun, Educational Testing Institute, Reykjavik, 2005), p. 26.

114 Ibid., 44.

International Survey (TALIS), TIMSS, and the national diagnostic tests.<sup>115</sup> In 2023, it was discovered that the Oslo department had analyzed the national test with a tool that produced wrong results in the period from 2014 to 2021.<sup>116</sup> The analyses of the development of the national test scores were then handed over to CEMO, and from 2022, new analyses tools were applied.<sup>117</sup>

Taking a step back, it is nevertheless true that, during the crucial years from 1954 to 1960, the Norwegian school authorities obviously used psychometric tests as an instrument to push through a major school reform. Here, the primary focus was on tests as tools for political and bureaucratic management rather than as instruments for pupils' growth and learning. However, pedagogical considerations about what would best serve the individual pupils always informed the development of the tests. The school maturity tests were intended to inform teachers, pupils, and parents at the start of schooling, and when later choosing educational tracks, in accordance with the individual pupil's abilities, interests, and skills. From the 1960s onwards, as we have seen, the tests in primary schools were used only as advisory tools for parents and pupils. Not to mention the crucial aspect of Sandven's research programme—to develop psychometric instruments to assess the pupils' emotional security and social attitudes, which resulted in serious critique of the Labor Party school reforms and never gained any political significance in the Norwegian school.

To summarize, it is fair to say that psychometric tests have played a significant role, both in the critical initial phase of developing the social democratic unitary school after World War Two, as well as in the international and strongly OECD-inspired development of the Norwegian school after PISA. The latter has in no way delivered the desired results, even though 2.6 billion NOK are now annually spent on operating the NQAS.<sup>118</sup> In recent years, the results from PISA, TIMSS, PIRLS, and ICCS have been declining. Over the twenty-year period, the

115 Institutt for lærerutdanning og skoleforskning, *Forskningsprosjekter fra A til Å*, Prosjekter - Institutt for lærerutdanning og skoleforskning (uio.no).

116 White paper to the Parliament, Nr. 34 (2023–2024), *En mer praktisk skole. Bedre læring, motivasjon og trivsel på 5.–10. trinn* (Ministry of Education and Research, 2024). p. 108.

117 Utdanningsdirektoratet, *Feil i resultater fra nasjonale prøver 2014–2021* (Utdanningsdirektoratet, 2024).

118 NOU 2023:1, 40.



progress has been stagnant at best. Regarding social studies education, the latest International Civic and Citizenship Education Study (ICCS) shows that Norwegian fourteen-year-olds' knowledge of democracy has declined the most among all participating countries from 2016 to 2022, and Norwegian students score significantly lower than their peers in Sweden and Denmark.<sup>119</sup> The leader of the latest evaluation of the Norwegian Quality Assessment System concludes that there is a need 'for a clear shift in thinking about quality assessment and quality development at all levels'.<sup>120</sup> She found strong support for her desire to improve the NQAS from two professors at CEMO—the then director Sigrid Blömeke and her colleague Rolf Vegar Olsen. In 2018, they wrote that the strictly nationally developed parts of the NQAS still lacked coherence and contained obvious contradictions, as it portrayed itself to be both a pedagogical tool and an instrument for management:

Even though the Norwegian school system conducts enough measurements, these individual assessments are disconnected from each other. The ambition should be to develop a quality system that offers a comprehensive analytical framework. In such a system, it should be possible to view the different assessments in relation to one another, with a stronger emphasis on a developmental perspective by facilitating the measurement of progression.<sup>121</sup>

This chapter also shows that the pupil-oriented pedagogical potential for empowerment, growth, and development, which was so deeply rooted in the largely forgotten psychometric legacy of Helga Eng and Johannes Sandven, has been scarcely explored during the latest and predominantly management-oriented psychometric wave in the making of the Norwegian school.

119 Oddveig Storstad, Joakim Caspersen and Christian Wendelborg, *Ett steg fram og to tilbake. Demokratiforståelse, holdninger og deltakelse blant norske ungdomsskoleelever* [The International Civic and Citizenship Education Study (ICCS) 2022] (NTNU Samfunnsforskning AS, 2023). See also Utdanningsdirektoratet, *Nedgang i norske ungdomsskoleelevers kunnskaper om demokrati og medborgerskap. ICCS 2022—rapport om medborgerskap* (Utdanningsdirektoratet, 2023).

120 Professor Tine Prøitz at seminar in *Forum for Critical Thinking and Critical Perspectives in School and Teacher Education* at Oslo Metropolitan University, 4.6.2024, 64 mins: Testar og målingar etter PISA - mål, hensikt og framtid for kvalitetsvurderingssystemet til skulen - FilMet, <https://film.oslomet.no/testar-og-malingar-etter-pisa-mal>

121 Sigrid Blömeke and Rolf Vegar Olsen, 'På vei mot et sammenhengende nasjonalt kvalitetsvurderingssystem', *Acta Didactica Norge*, 12(4) (2018), 17–18, <https://doi.org/10.5617/adno.6278>

## Bibliography

- Allerup, Peter, Velibor Kovac, Gro Kvåle, Gjert Langfeldt and Poul Skov, *Evaluering av det Nasjonale kvalitetsvurderingssystemet for grunnsopplæringen* (FOU-rapport 8/2009, Agderforskning and Aarhus: The Danish School of Education, Aarhus University for The Norwegian Directorate for Education and Training, 2009).
- Blair, Tony and Gerhard Schröder, *Europe: The Third Way/Die Neue Mitte*. Working Documents (Friedrich Ebert Foundation, South Africa Office, 1998).
- Blömeke, Sigrid and Rolf Vegar Olsen, 'På vei mot et sammenhengende nasjonalt kvalitetsvurderingssystem', *Acta Didactica Norge*, 12.4 (2018), <https://doi.org/10.5617/adno.6278>
- Brehony, Kevin J., 'A new education for a new era: the contribution of the conferences of the New Education Fellowship to the disciplinary field of education 1921–1938', *Paedagogica Historica*, 40.5–6 (2004), 733–755, <https://doi.org/10.1080/0030923042000293742>
- Børhaug, Kjetil and Kim Helsvig, 'Kampen om samfunnsfaget—eit forskingsoversyn med vekt på lærarane si rolle', *Nordidactica—Journal of Humanities and Social Science Education* (2023).
- Dale, Erling Lars, *De strategiske pedagoger—Pedagogikkens vitenskapshistorie i Norge* (Ad Notam Gyldendal, 1999).
- Dewey, John, *Democracy and Education* (New York: The Macmillan Company, The Free Press, 1946).
- Dokka, Hans-Jørgen, *Reformarbeid i norsk skole* (NKS-Forlaget: 1986).
- Eide, Kjell, *Departementets lille kanarifugl, eller kulturpolitikk bli til* (NIFU, 1985).
- Eide, Kjell, *OECD og norsk utdanningspolitikk. En studie av internasjonalt samspill* (NIFU, 1995).
- Eidsvik, Erlend, 'Fagfornying, berekraft og geografisk danning', *Norsk pedagogisk tidsskrift*, 104.3 (2020), 268–282.
- Hareide, Bjarne 'Focus' *Prismet* 8/1950 (IKO-forlaget, 1950).
- Helsvig, Kim, *Pedagogikkens grenser. Kampen om norsk pedagogikk ved Pedagogisk forskningsinstitutt 1938–1980* (Oslo: Abstrakt forlag, 2005).
- Helsvig, Kim, *Reform og Rutine. Kunnskapsdepartementets historie 1945–2017* (Pax Forlag, 2017).
- Helsvig, Kim, 'Kunnskapssamfunnet og nasjonen—den internasjonale vendingen', *Nytt Norsk Tidsskrift*, 39.1 (2022), <https://doi.org/10.18261/nnt.39.1>
- Husén, Torsten, *En obotlig akademiker—En professors memoarer* (Stockholm, 1981).

- Institutt for lærerutdanning og skoleforskning, *Forskningsprosjekter fra A til Å, Prosjekter - Institutt for lærerutdanning og skoleforskning* (uio.no)
- Jordheim, Knut, 'Okkupasjonsårene 1940–45 som "epoke" i norsk skoles utvikling' in *Skolen 1999–2000—Årbok for norsk utdanningshistorie*, ed. by Knut Jordheim (Stiftelsen SKOLEN—Årbok for norsk utdanningshistorie, 2000), pp. 58–76.
- Karlsen, Gustav E., 'Internasjonale aktører som premissgivere for norsk utdanningspolitikk med særlig vekt på OECD', in *Reformideer i norsk skole: Spredning, oversettelse og implementering*, ed. by K. A. Røvik, T.V. Eilertsen and E.M. Furu (Cappelen Damm akademisk, 2014).
- Kolsrud, Ole, *Rekonstruksjon og reform. Regjeringskontorene 1945–2005* (Universitetsforlaget, 2008).
- Kunnskapsdepartementets avdeling for analyse-, internasjonalt arbeid og kompetansepolitikk, AIK, *Kunnskapsdepartementets strategi for utdanningsforskning. Kunnskap for kvalitet 2008–2013* (Kunnskapsdepartementet, 2008).
- Langfeldt, Gjert and Torunn Lauvdal, *Evaluerings av det nasjonale kvalitetsvurderingssystemet. Delrapport pr. 1.4.2006* (The Norwegian Directorate for Education and Training, 2006).
- Lie, Svein, Marion Caspersen and Julius K. Björnsson, *Nasjonale prøver på prøve. Rapport fra en utvalgsundersøkelse for å analysere og vurdere kvaliteten på oppgaver og resultater til nasjonale prøver våren 2004* (Department of Teacher Education and School Research, University of Oslo and Námssmáttstofnun, Educational Testing Institute, Reykjavik, 2005).
- Lønnå, Elisabeth, *Helga Eng—Psykolog og pedagog i barnets århundre* (Fagbokforlaget, 2002).
- Marklund, Sixten, *Skolesverige 1950–1975. Del 1—1950 års reformbeslut* (Skolöverstyrelsen och Liber Utbildningsförlaget, 1980).
- McKnight, A., H. Glennerster, and R. Lupton, 'Education, education, education ... an assessment of Labour's success in tackling education inequalities', in *A More Equal Society?: New Labour, Poverty, Inequality and Exclusion*, ed. by J. Hills and K. Stewart (Bristol University Press, 2005), pp. 47–68.
- Mehta, Jal, 'Escaping the Shadow: A Nation at Risk and Its Far-Reaching Influence', *American Educator*, 39.2 (2015), 20–26.
- Melding om forsøk i skolen 1959–60, *Norsk skole. Opplysnings- og kunngjøringsblad for skoleverket*, nr. 6 (1962).
- Nordin, A. and D. Sundberg, eds., *Transnational Policy Flows in European Education: The Making and Governing of Knowledge in the Education Policy Field* (Oxford: Symposium Books, 2014).
- Norwegian government (2000), *Regjeringen Stoltenbergs tiltredelseserklæring - regjeringen.no*.

- Norwegian government (2001), Regjeringen Kjell Magne Bondeviks tiltredelseserklæring - regjeringen.no.
- NOU 2003: 16. *I første rekke—Forsterket kvalitet i en grunnopplæring for alle*, NOU 2003: 16 - regjeringen.no.
- NOU 2023: 1. *Kvalitetsvurdering og kvalitetsutvikling i skolen. Et kunnskapsgrunnlag*, NOU 2023: 1 - regjeringen.no.
- NOU 2023: 27. *Et nytt system for kvalitetsutvikling—for elevenes læring og trivsel*, NOU 2023: 27 - regjeringen.no.
- OECD, *The Knowledge-Based Economy* (OCDE/GD, 1996).
- OECD, *OECD Economic Surveys: Norway 1997* (OECD Publishing, 1997).
- OECD, *Human Capital Investment: An International Comparison* (OECD Publishing, 1998).
- OECD, *The Well-being of Nations: The Role of Human and Social Capital* (OECD Publishing, 2001).
- OECD, *Science Competencies for Tomorrow's World Volume 1* (Paris: Organisation for Economic Co-operation and Development, 2007).
- OECD, *The High Cost of Low Educational Performance: The Long-run Economic Impact of Improving PISA Outcomes* (PISA, OECD Publishing, 2010).
- OECD, *PISA 2015 Results (Vol. II): Policies and Practices for Successful Schools* (OECD Publishing, 2016).
- Oelkers, Jürgen, 'Origin and Development in Central Europe' in *Progressive Education Across the Continents*, ed. by Hermann Röhrs and Volker Lenhart (Peter Lang, 1995), pp. 31–50.
- Papadoupoulos, George S., *Education 1960–1990—The OECD Perspective* (OECD, 1994).
- Pettersson, Daniel and Thomas S. Popkewitz, 'A chimera of quantifications and comparisons. The changing of educational "expertise"', in *New Practices of Comparison, Quantification and Expertise in Education*, ed. by Christina Elde Mølstad and Daniel Pettersson (London: Routledge, 2019), <https://doi.org/10.4324/9780429464904-2>
- Pokropek, Artur, Gary N. Marks, Francesca Borgonovi, Piotr Koc and Samuel Greiff, 'General or specific abilities? Evidence from 33 countries participating in the PISA assessments', *Intelligence*, 92 (2022), <https://doi.org/10.1016/j.intell.2022.101653>
- Popkewitz, Thomas S., 'Infrastructures and phantasmagrams of inclusions that exclude: international student assessments', *International Journal of Inclusive Education*, 29.8 (2023), 1424–1444, <https://doi.org/10.1080/13603116.2023.2275156>

- Rand, Per, 'Hovedtrekk ved Johs. Sandvens forskningsinnsats', in *Pedagogikkens søkelys. Festskrift til professor Johs. Sandven på 70-årsdagen 1. november 1979* (Universitetsforlaget, 1979), pp. 1–18.
- Röhrs, Hermann, 'Internationalism in Progressive Education and Initial Steps towards a World Education Movement', in *Progressive Education Across the Continents*, ed. by Hermann Röhrs and Volker Lenhart (Peter Lang, 1995), pp. 11–27.
- Røvik, Kjell Arne, Tor Vidar Eilertsen and Torbjørn Lund, 'Hvor har de det fra, og hva gjør de med det? Utdanningsdirektoratet som innhenter, oversetter og iverksetter av reformideer', in *Reformideer i norsk skole Spredning, oversettelse og implementering*, ed. by Kjell Arne Røvik, Tor Vidar Eilertsen and Eli Moksnes Furu (Oslo: Cappelen Damm, 2014), pp. 88–118.
- Sandven, Johs., *Pedagogisk idébrytning i U.S.A.—En studie over utviklingen og de rådende synsmåter i dag* (Oslo: Fabritius og Sønners Forlag, 1949).
- Sandven, Johs., *Målsetningen i folkeskolen*, Special edition of *Skole og samfunn*, 3 (1953).
- Sandven, Johs, 'Fremtidens skole under pedagogisk-psykologisk synsvinkel', Special edition of *Pedagogisk forskning, Nordisk tidsskrift for pedagogikk*, 3 (1958).
- Sandven, Johs, 'Det teoretiske og metodiske grunnlag for modenhetsprøving: De norske modenhetsprøver for barne- og ungdomsalder', *Pedagogisk Forskning*, 6.1 (1962), 147–168, <https://doi.org/10.1080/0031383620060110>
- Sandven, Johs, 'Årsaker til manglende skoletrivsel', *Forskning med relation til utbildning för åldersklasserna 16–19 år* (Nordisk Komité for Pedagogisk Forskning, 1971).
- Seitzer, Helen, D. Niemann and K. Martens, 'Placing PISA in perspective: the OECD's multi-centric view on education', *Globalisation, Societies and Education*, 19.2 (2021), 198–212, <https://doi.org/10.1080/14767724.2021.1878017>
- Sivertsen, Helge, 'Norsk skolereform', *Forsøk og reform* (Forsøksrådet for skoleverket, 1956).
- Sirevåg, Tønnes, *Den ni-årige skolen* (Aschehoug, 1979).
- Sjøberg, S., 'PISA-syndromet—Hvordan norsk skolepolitikk blir styrt av OECD', *Nytt Norsk Tidsskrift*, 31.1 (2014), 30–43.
- Skjervheim, Hans, 'Den eigne refleksjon og fortolkinga av andre', in *Deltakar og tilskodar* (Tanum-Norli, 1976), pp. 287–299.
- Skjervheim, Hans, Eit grunnproblem i pedagogisk filosofi, 226–245 in *Deltakar og tilskodar* (Tanum-Norli, 1976b).
- Skjervheim, Hans and Leif Tufte, eds., *Pedagogikk og samfunn. Foredrag på Vosse-seminaret 13.-15.-desember 1968* (Gyldendal, 1969).
- Stafseng, Ola, 'Helga Eng og hennes pedagogikkfag 1938—mellom tradisjon og arvefølge', *Norsk Pedagogisk Tidsskrift*, 93.6 (2009), <https://doi.org/10.18261/ISSN1504-2987-2009-06-03>

- Storstad, Oddveig, Joakim Caspersen and Christian Wendelborg, *Ett steg fram og to tilbake. Demokratiforståelse, holdninger og deltakelse blant norske ungdomsskoleelever* [The International Civic and Citizenship Education Study (ICCS) 2022] (NTNU Samfunnsforskning AS, 2023).
- Sørensen, Tore Bernt, Christian Ydesen and Susan Lee Robertson, 'Re-reading the OECD and education: the emergence of a global governing complex—an introduction, 101' *Globalisation, Societies and Education*, 19.2 (2021), <https://doi.org/10.1080/14767724.2021.1897946>
- Telhaug, Alfred Oftedal, *Forsøksrådet for skoleverket (1954–1984)* (Universitetsforlaget, 1990).
- Telhaug, Alfred Oftedal, A.I. Wallum and R.T. Tønnesen, *Norsk utdanningspolitisk retorikk, 1945–2000* (Cappelen Akademisk, 1999).
- Thue, Fredrik and Kim Helsvig, *Universitetet i Oslo. Den store transformasjonen* (Oslo: Unipub, 2011).
- Ubøe, Jan, 'PISA har utspilt sin rolle', *NHH-Bulletin* (2016), PISA har utspilt sin rolle | NHH. <https://www.nhh.no/nhh-bulletin/artikkelarkiv/eldre-saker/2016/desember/pisa-har-utspilt-sin-rolle/>
- Utdanningsdirektoratet, *Delrapport, Nasjonalt kvalitetsvurderingssystem* (Utdanningsdirektoratet, 2010).
- Utdanningsdirektoratet, *Nedgang i norske ungdomsskoleelevers kunnskaper om demokrati og medborgerskap. ICCS 2022—rapport om medborgerskap* (Utdanningsdirektoratet, 2023).
- ICCS 2022—rapport om medborgerskap | udir.no. <https://www.udir.no/tall-og-forskning/finn-forskning/rapporter/2023/iccs-2022-rapport-om-medborgerskap/>
- Utdanningsdirektoratet, *Feil i resultater fra nasjonale prøver 2014–2021* (Utdanningsdirektoratet, 2024), *Feil i resultater fra nasjonale prøver 2014–2021* | udir.no. <https://www.udir.no/tall-og-forskning/feil-i-resultater-fra-nasjonale-prover-20142021/>
- White paper to the Parliament, *Nr. 28 (1998–99), Mot rikare mål* (1999), *St.meld. nr. 28 (1998-99)* - regjeringen.no
- White paper to the Parliament, *Nr. 22 (2010–2011), Motivasjon—Mestring—Muligheter* (Ministry of Education and Research, 2011), *Meld. St. 22 (2010 2011)* - regjeringen.no
- White paper to the Parliament, *Nr. 34 (2023–2024), En mer praktisk skole. Bedre læring, motivasjon og trivsel på 5.–10. trinn* (Ministry of Education and Research, 2024), *Meld. St. 34 (2023–2024)*
- Ydesen, Christian and Karen Egedal Andreasen, 'Historical Roots of the Global Testing Culture in Education', *Nordic Studies in Education*, 40.2 (2020), <https://doi.org/10.23865/nse.v40.2229>



# 13. Drawing Boundaries, Building Barriers: Twentieth-Century US and Norwegian Intelligence Testing from a Comparative Perspective

*Jim Wynter Porter*

---

This chapter sits as a transatlantic companion to this volume's Scandinavian histories. Its aim is to survey the US history of IQ alongside the Norwegian accounts, and derive a comparative analysis focused on the use of IQ in educational contexts. There are good reasons why the US might be a generative comparative foil for Norway in this regard. The US was a hotbed of early development and transnational exchange in "intelligence" testing, and as John Carson observes, exceptional among nations in the degree to which IQ testing expanded within educational settings.<sup>1</sup> Yet, while most other countries abandoned efforts to formalize intelligence testing early in the twentieth century, Norway (and other Scandinavian nations) would institutionalize IQ for particular educational functions, and thus follow some distance down the path the US had taken.

Given the large difference in size between the US and Norway, my collaborators on this volume have at times remarked that putting them next to one another is a little like comparing an elephant and a mouse. In 1920, the US had about a hundred million people spread across its forty-eight states. Norway, slightly smaller in square kilometers than the

---

1 John Carson, 'Mental Testing in the Early Twentieth Century: Internationalizing the Mental Testing Story', *History of Psychology*, 17.3 (2014), 249–255, (pp. 253–54), <https://doi.org/10.1037/a0037475>



state of California, was then home to around 2.5 million people. The reader should bear these differences in scale in mind as it undoubtedly influenced the dynamics considered below.

Much of this chapter explores the US history of IQ testing with a particular focus on three mutually reinforcing factors that help explain similarities and differences between US and Norwegian practice across the twentieth century: 1) racism and classism, and the extent to which these systemic ideologies shaped beliefs about the educability of various social groups; 2) educational policy developments and the extent to which they structured differentiation in schools; and 3) market forces and the extent to which they incentivized testing and in turn shaped policy. As this chapter hopes to show, the degree to which these factors acted together and reinforced one another help explain the speed and scope of IQ's expansion in both national cases. In retelling the US story through this particular lens, this chapter relies on the excellent prior historical analyses of John Carson, Paul Davis Chapman, and Leila Zenderland.<sup>2</sup>

Carson and others have identified additional factors important to the story of US expansion, including: 4) the onset of the government-funded World War One US Army testing programme, which served as a catalyst for mass testing; and 5) the large size of the US psychological profession in the early twentieth century, and the greater degree of professional status staked on IQ there.<sup>3</sup> Notably, psychology did not professionalize in Norway until after World War Two. Instead, psychiatry and pedagogy were the initial homes of Norwegian IQ testing and together constituted a smaller body of experts than US psychologists.<sup>4</sup> These developments certainly played a major role in shaping IQ's expansion, and helped

2 John Carson, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750–1940* (Princeton: Princeton University Press, 2007); Paul Davis Chapman, *Schools as Sorters: Lewis M. Terman, Applied Psychology, and the Intelligence Testing Movement, 1890–1930* (New York University Press, 1990); Leila Zenderland, *Measuring Minds: Henry Herbert Goddard and the Origins of American Intelligence Testing* (Cambridge: Cambridge University Press, 2001).

3 Carson, 'Mental Testing in the Early Twentieth Century', p. 254.

4 Lina Bryne Larsen, 'Med måling følger viden» Intelligensmåling i den norske folkeskolen 1900–1945' (Master's thesis, Historisk Institutt, University of Bergen, 2007), p. 66; Emberland, chapter 5, this volume; Lefkadiou, chapter 1, this volume; see also Ydesen, Ludvigsen, and Lundahl, 'Creating an Educational Testing Profession in Norway, Sweden and Denmark, 1910–1960', p. 128; Kari Ludvigsen and Åsmund Arup Seip, 'The Establishing of Norwegian Child Psychiatry: Ideas, Pioneers and Institutions', *History of Psychiatry*, 20.1 (1 March 2009), 5–26, <https://doi.org/10.1177/0957154X08089847>

account for US-Norwegian divergence. This chapter, however, chooses to devote its attention to the trio identified above—markets, policy, and systematic prejudice—as the interactions between these factors appears to also have high explanatory potential.

There is finally an additional underlying ideology—IQ hereditarianism—that sustained these mutually reinforcing factors. A misconception garbed in the trappings of science, hereditarian IQ theory held that IQ variance must correspond to putative natural, inherent differences among learners. Without this stabilizing (and “naturalizing”) belief system in place, policy and market forces which cooperated so readily with racism and classism in the US case might otherwise have more easily come undone.

## Background and Overview

The US account presented here will follow testing from its beginnings in special education through two waves of expansion across the general student population, and it will be complemented by reflections on Norwegian developments across the same time periods. But first, let me briefly frame both the central commonality and difference between US and Norwegian testing practice. Both Norway and the US would initially adopt individual IQ testing (one tester—one examinee) for the purpose of identifying so-called “feeble-minded” or *svakt begavede* [weakly gifted] children, either for institutionalization or for assignment to special education classes. Yet, while Norwegian practice would largely be restrained here, the US would carry IQ testing further, and do so more rapidly. By 1920, with the development of new group tests, and the creation of a conducive policy context, American IQ testing would expand as mass practice, and not only for identification for special education, but also for “ability” grouping and tracking across the general school-aged population.<sup>5</sup> By 1958, again propelled by new US policy, such mass testing would undergo a second wave of more comprehensive expansion.

There are a few exceptions to this general framing on the Norwegian side of the ledger. Firstly, from 1934 into the mid-1950s, Oslo schools—

---

5 John Carson, *The Measure of Merit*, pp. 159, 179, 218–219; Paul Davis Chapman, *Schools as Sorters*, pp. 22–28, 42; Leila Zenderland, *Measuring Minds*, pp. 257–264, 281–290.

and schools in some other large city systems—also used group tests, but did so predominantly for special education screening, and not for tracking. Yet, it is also clear that in this period there were incipient efforts to use group tests to assess general ability in some large Norwegian municipalities, again especially Oslo. Notably, though, such early general ability testing was less systematic than in the US, even some twenty years later. Finally, in the mid-1950s, Norwegian policy makers did consider—and experiment with—more systematic group testing to sort the general school-aged population by ability, but this project was soon abandoned.<sup>6</sup>

What early conditions establish a basic starting point for national testing practices in both countries? Let us consider some contextual overlap first. Both the US and Norway were engaged in their own early twentieth-century modernization projects, which sought to extend public schooling to greater portions of their nation's children. Such expansions in both cases prompted increasing visibility of educational differences in the classroom. This, in turn, impelled attempts to define “normality” and “deficiency,” ideas which would soon regulate full inclusion in the new modern school.<sup>7</sup> So-called intelligence tests became a primary mechanism for such regulation in the US and Norway, and were embraced in both cases by progressive reformers who believed them to be crucial to such a project.<sup>8</sup>

As many of the contributors to this volume observe, there was also ample transnational exchange of tests: first from France to both countries, and then later from the US to Norway. Along with this exchange came similar problems concerning the adaptation of these instruments for diverse national populations. Both the US and Norway embarked on testing projects that never adequately accounted for the degree to which

6 Lina Bryne Larsen, *Med måling følger viden*, pp. 50–51; Lefkaditou, chapter 1, Helsvig, chapter 12, and Thue, chapter 11, this volume.

7 As shown in different instantiations in Norway by, for instance, Thue, chapter 11, this volume; Ydesen, Hovland and Vikstrom, chapter 2, this volume; and Lefkaditou, chapter 1, this volume.

8 Christian Ydesen, Kari Ludvigsen, and Christian Lundahl, ‘Creating an Educational Testing Profession in Norway, Sweden and Denmark, 1910–1960’, *European Educational Research Journal*, 12.1 (1 March 2013), 120–138 (pp. 121, 125), <https://doi.org/10.2304/eerj.2013.12.1.120>; Alfred Oftedal Telhaug, Odd Asbjørn Mediås, and Petter Aasen, ‘The Nordic Model in Education: Education as Part of the Political System in the Last 50 Years’, *Scandinavian Journal of Educational Research*, 50.3 (1 July 2006), 245–283 (p. 249), <https://doi.org/10.1080/00313830600743274>

tests were biased by ethnicity and class.<sup>9</sup> Proponents of testing in both countries continually sidestepped this problem, and instead generally promoted IQ as an objective measure of inherent ability.

Issues of racism and classism—so often central to histories of intelligence testing—invite careful comparative scrutiny here. Both US and Norwegian national majorities engaged in the racialization of their minority populations, but the aims of these racial projects and the extent to which they were institutionalized vary in both cases. The US has long been more ethnically diverse than Norway, but this rich diversity grows out of a deeply troubled history, marked by the genocide, removal, and relocation of indigenous populations, and by slavery—a brutal labor regime which evolved, post-emancipation, into an economic system that remained dependent on the exploitation of non-white and lower-class labor.<sup>10</sup> Norway was not a slavery-dependent state.<sup>11</sup> Though it also, like the US, had a eugenics movement, Norway's racial project is better historicized as assimilationist. National minorities, such as the Sámi and Kvens, while identified as racially “other”, were generally subjected to ‘Norwegianization’, a process that depended intensively on their ‘re-education’ and adoption of majority language and cultural practices.<sup>12</sup>

Such early differences in national histories of race and class are important to this comparative analysis. They have, for one, given rise over the twentieth and early twenty-first centuries to distinct versions of capitalism in each nation. US free market capitalism, with its relatively vulnerable labor movement and social welfare system, contrasts decidedly with Norway's social democratic welfare state and politically consequential Labor party. These distinct political-economic trajectories—and the relative degrees of exploitation or solidarity they engender—shape the history of IQ testing at almost every level.

9 See for instance Kyllingstad, chapter 10, this volume; Paulsen, chapter 9, this volume; and Lefkaditou, chapter 1, this volume.

10 Matthew Desmond, ‘Capitalism’, in *The 1619 Project: A New Origin Story*, ed. Nikole Hannah-Jones and The New York Times Magazine (Random House Publishing Group, 2021).

11 Norwegians were involved in the Atlantic slave trade and the Danish-Norwegian slave-based overseas colonies in the eighteenth and early nineteenth century. But there was no race-based slavery on Norwegian territory, and there is today no racialized population of descendants of slaves in Norway.

12 Jon Røyne Kyllingstad, *Measuring the Master Race: Physical Anthropology in Norway, 1890–1945* (Open Book Publishers, 2014); Henry Minde, ‘Assimilation of the Sami - Implementation and Consequences’, *Journal of Indigenous Peoples Rights*, 3 (1 January 2005).

## Vineland and the Institutional Origins of US “Intelligence” Testing

The modern intelligence test was invented in France in 1905 by psychologist Alfred Binet and psychiatrist Théodore Simon. Binet had been commissioned by the French Ministry of Public Instruction to create an instrument that could identify ‘*anormaux*’ students who lagged significantly behind their peers in school progress. The test they developed yielded a ‘mental age’ score which compared a student’s performance in relation to their age cohort. Because its questions were selected by gauging the responses of about 200 “normal” Parisian children, the finalized scale could be said to make this age comparison more consistently and reliably than any regular pedagogical test.<sup>13</sup> American psychologist Henry Goddard obtained a copy of the Binet–Simon intelligence test while traveling in France. Goddard worked as a research psychologist at the Vineland Training School for Boys and Girls, a residential institute for so-called “feeble-minded” children and young adults in New Jersey. Goddard wanted to establish a science of “feeble-mindedness”, and came to believe that this new test held the key.<sup>14</sup>

What was feeble-mindedness? No one really knew. At the turn of the century, this alleged condition was a medical diagnosis based on a highly subjective assessment that could vary greatly from doctor to doctor, and which depended on the evaluation of any number of physical traits then thought to betoken the condition.<sup>15</sup> Goddard was unsatisfied with current medical approaches, and after translating—but not norming—the 1908 revision of the Binet–Simon Test of Intelligence, he began to use it experimentally with residents at Vineland.<sup>16</sup>

It is clear in hindsight that Goddard’s young wards had learning difficulties arising from a wide variety of issues. Some had sensory or speech impairments. Some had been debilitated by early childhood disease. Other residents had been orphaned, or had been exposed to extreme poverty, abuse, or trauma, or lacked access to formal education.

13 John Carson, *The Measure of Merit*, pp. 139–43; Leila Zenderland, *Measuring Minds*, pp. 93–94, 96; Walter Lippmann, ‘The Mental Age of Americans’, *New Republic*, 32.412 (25 October 1922), 213–215 (p. 213).

14 Zenderland, *Measuring Minds*, p. 70.

15 *Ibid.*, p. 75.

16 *Ibid.*, pp. 95–102.

Some had congenital conditions like Down's syndrome (then called 'mongoloidism'), or suffered from neurological issues such as cerebral palsy or epilepsy.<sup>17</sup> Yet, in Goddard's day, most of these issues were commonly heaped together—simplistically and inaccurately—as manifestations of feeble-mindedness.

When Goddard applied this new intelligence test to Vineland children, he found that the Binet-Simon agreed well with his own evaluations of degrees of educational deficiency, resident by resident.<sup>18</sup> The scale was, with an appreciable degree of reliability, telling Goddard where a child stood in relation to his or her age cohort. Even so, this 'mental age' score itself said nothing about *why* these measured differences might have arisen in the first place, nor how fixed or stable they were. Nor did it say anything about whether the 'norm' against which these children were compared—what the test presumed a person of a given age should know—had been fairly arrived at. Yet, because this approach yielded reproducible measurements, Goddard believed he was a big step closer to clarifying what this 'condition' was, how it should be dealt with, and even what caused it.

Goddard now redefined feeble-mindedness, based on the Binet-Simon, as a person having a mental age more than three years behind their chronological age.<sup>19</sup> He argued that his approach was superior to medical diagnostics, noting the reliability of this measure, and that the Binet-Simon was a test of actual mental performances, not an assessment of mentality by proxy, approached through a farrago of supposed physio-medical signifiers. As Goddard understood it, feeble-mindedness was now a scientifically objective, measurable 'deficiency in intelligence'.<sup>20</sup>

The consistency of the results, combined with his hereditarian inclinations, led Goddard to suspect that this alleged condition was also preponderantly, or even exclusively, hereditary in origin.<sup>21</sup> In his attempt to thus forge a hard science of mental deficiency, he would continue to inadequately rule out or fail to consider the role social and environmental factors played in the diverse learning issues he was

---

17 Ibid., pp. 76–79.

18 Carson, *The Measure of Merit*, p. 178.

19 Zenderland, *Measuring Minds*, p. 121.

20 Ibid., p. 104.

21 Ibid., pp. 28–39, 85.

trying to understand.<sup>22</sup> Goddard himself was an energetic promulgator of this new method of diagnosis, and word spread quickly among US psychologists and educators. Teachers came to Vineland to learn how to administer this new intelligence test, and by 1914, Goddard's translation of the Binet–Simon was in regular use in hundreds of schools.<sup>23</sup> By 1916, he had mailed 22,000 copies of the test to teachers around the country.<sup>24</sup>

### Education for the Masses, Schooling on the Cheap: The “Feeble-minded” and “Laggards” in American Public Schools

The USA underwent massive population growth from the 1880s through the 1920s, augmented by a steep rise in immigration, particularly of people from Southern and Eastern Europe. This was coupled with increased urbanization, and the migration of African Americans who fled the Jim Crow South for the more industrial areas of the North and West. Whites were sometimes no longer a majority among student bodies, especially in cities.<sup>25</sup> The general growth and diversification of the American population was compounded in schools with the enactment of new compulsory education laws which, by 1920, typically mandated school attendance until sixteen years of age.

Enrolments thus skyrocketed around the country, especially in high schools and in urban areas. Between 1890 and 1915, average national elementary school enrolments rose 47 percent, and the high school attending population vaulted by a staggering 554 percent.<sup>26</sup> Even in better-resourced districts, classrooms might be crammed with over seventy students.<sup>27</sup> It was likewise increasingly apparent that many students in major school systems around the country were failing school or lagging appreciably behind.<sup>28</sup>

---

22 Ibid., p. 124.

23 Ibid., pp. 114–15, 137, 138.

24 Ibid., p. 141.

25 Jeannie Oakes, *Keeping Track: How Schools Structure Inequality* (Yale University Press, 1986), p. 20.

26 Chapman, *Schools as Sorters*, p. 42.

27 Marvin Lazerson, *Origins of the Urban School: Public Education in Massachusetts, 1870–1915* (Harvard University Press, 2013), pp. 11–12.

28 Zenderland, *Measuring Minds*, pp. 105–108.

Such conditions were exacerbated by conservative fiscal politics generally resistant to centralized funding of public education.<sup>29</sup> Instead, US schools were (and continue to be) financed primarily through local tax revenue, a mechanism that reproduced sharp educational inequalities across communities: wealthy, middle-class, and poor. US public schools had to respond to an increasing number and diversity of students—by culture, class, language, general school-readiness, and learning differences—without corresponding increases in centralized funding or resources.

It was within this set of constraints that Progressive-era officials and school administrators—favoring rational-technocratic solutions to social problems—sought to make schools more efficient through principles of ‘scientific management’. Several ensuing school efficiency studies attracted significant public attention and helped redefine the current challenges of public schools as not an under-resourced educational system, but a problem resulting from large numbers of individuals not making expected progress with the standard curriculum, namely ‘laggards’ and ‘over-aged’ pupils.<sup>30</sup>

These emergent debates gave Goddard an opportunity to bring the Binet–Simon into public schools on a large-scale experimental basis to assess how many of these laggards might actually be, by his measure, feeble-minded. Between 1910 and 1912, Goddard tested thousands of students in New Jersey and New York City and, after some extrapolations, sensationally publicized that New York City public schools had six times more undiagnosed feeble-minded students than were currently enrolled in special classes. The resulting inefficiency, it was argued, could be rectified if feeble-minded children were more consistently “diagnosed” and segregated.<sup>31</sup>

---

29 Carl Kaestle and Marshall Smith, ‘The Federal Role in Elementary and Secondary Education, 1940–1980’, *Harvard Educational Review*, 52, no. 4 (1982), 384–408.

30 Chapman, *Schools as Sorters*, p. 44.

31 Zenderland, *Measuring Minds*, pp. 120–34.



## The Nature of Feeble-mindedness and Hereditarian Roots of Intelligence

Goddard's early approach to intelligence testing was also forging strong continuities with the eugenics movement. Eugenics was a hereditarian pseudo-science—organized around prejudicial perceptions of race, class and human “quality”—that sought the biological improvement of society. In tandem with this “scientific” mission, eugenics was also a political movement that lobbied for population management and reproductive control through such means as immigration restriction and involuntary sterilization.<sup>32</sup> American eugenicists became enchanted with intelligence testing as a supposedly objective means for marking out natural distinctions within a rapidly changing social order.<sup>33</sup> At the urging of eugenicist and geneticist Charles Davenport, Goddard undertook two major studies—*The Kallikak Family* (1912), and *Feeble-mindedness: Its Causes and Consequences* (1914)—explicitly investigating the putative hereditary origins of feeble-mindedness. Eugenic “field workers” were sent into some of the poorest communities in New Jersey to excavate extended family histories of Vineland children. The method was interview-based and predicated on the belief that field workers who had trained with the Binet–Simon could simply spot a feeble-minded person when they talked to one, or even heard accounts of individuals amidst garbled compilations of family history and gossip. Based on such “data”, Goddard and his field workers then reconstructed normal/feeble-minded Vineland pedigrees running back nearly nine generations, from the early 1700s to around 1900.<sup>34</sup>

Davenport convinced Goddard to apply the new Mendelian theory of inheritance to the results of these studies. Based on the frequencies of occurrence Goddard believed he had documented, he now proclaimed that feeble-mindedness was the result of a single Mendelian unit character,

---

32 Carson, *The Measure of Merit*, p. 160; Daniel J. Kevles, *In the Name of Eugenics: Genetics and the Uses of Human Heredity* (University of California Press, 1995); Diane B. Paul, *Controlling Human Heredity* (Humanity Books, 1999).

33 Carson, *The Measure of Merit*, p. 191.

34 Stephen Jay Gould, *The Mismeasure of Man* (New York: W. W. Norton & Company, 2006), pp. 198–203; Zenderland, *Measuring Minds*, pp. 158–162, 172.

or “gene”.<sup>35</sup> Though later scientists and historians would debunk Goddard’s claims in these works, *Kallikaks* and *Feeble-mindedness* were received by their contemporaries as cutting-edge science, and widely read and understood as a discovery of facts about human nature.<sup>36</sup> *The Kallikaks* became one of the most popular publications associated with the eugenics movement.<sup>37</sup>

As Goddard began his public-school surveys, he was likewise invited by the Public Health Service to review immigrant screening procedures.<sup>38</sup> Goddard himself was not a proponent of immigration restriction, yet after extensive Binet testing at Ellis Island, he would report a shocking forty percent rate of feeble-mindedness among incoming Southern and Eastern European immigrants. This was nearly twenty times his estimate for the general US-born population. Such wildly aberrant results prompted Goddard to question his own methods, and here he began to consider the possible role of education, culture, and linguistic difference on his results.<sup>39</sup> Yet, despite his reservations, his rates of immigrant feeble-mindedness, published in 1917, were readily promulgated by a mounting immigration restriction movement.<sup>40</sup>

### Beyond Feeble-mindedness: IQ Testing for the General School-Aged Population

Already by 1914, intelligence had become one of psychology’s central concepts, and one of its most reliably reproducible measurements.<sup>41</sup> The Binet–Simon was ever more widely promoted in professional journals for psychology, and there was a growing tendency to operationalize intelligence as a continuous variable, which might be measured by degrees and across the general population.<sup>42</sup> This broadening conception of intelligence was likewise widely understood (though not

---

35 Ibid., pp. 171, 189–91.

36 Ibid., p. 144.

37 Ibid., p. 158, 169.

38 Ibid., pp. 266–68.

39 H.H. Goddard, ‘Mental Tests and the Immigrant’, *Journal of Delinquency* 2 (1917), 243–277 (pp. 245–51, 270).

40 Zenderland, *Measuring Minds*, pp. 279, 268–280.

41 Carson, *The Measure of Merit*, p. 159.

42 Ibid., p. 179.

without debate and disagreement) as a measure of natural, hereditary superiority and inferiority of both groups and individuals. Such expansions of intelligence included early use of tests by psychologists to differentiate among pupils performing at or above grade-level in schools.<sup>43</sup> US schools themselves, overenrolled and saddled with calls for greater efficiency, were increasingly interested not only in separating out those who lagged behind, but in sorting all their students.<sup>44</sup> Recall, this eventual development—the establishment of testing across the general student population for ability grouping and tracking—marks perhaps the clearest and most significant difference between intelligence testing practice in the US and Norway.

Chief among these early practitioners of school-place intelligence testing was Stanford University psychologist Lewis Terman. If Goddard was focused on the diagnosis of so-called feeble-mindedness, Terman was more interested in classifying intelligence across “normal” and “supra-normal” domains.<sup>45</sup> In 1916, Terman launched the new Stanford–Binet Test of Intelligence, now at last normed for a US population, for a wider range of ages—five- to sixteen-year-olds—and based on a much larger sample. Terman also employed a calculation devised by German psychologist William Stern—mental age divided by chronological age, multiplied by one hundred—to even out discrepancies at upper and lower ends of the age span. This intelligence quotient (or IQ), coupled with Terman’s new test, yielded a bell curve distribution with an average score of one hundred. Historian and psychologist Raymond Fancher notes that the Stanford–Binet quickly ‘became the standard against which all subsequent American intelligence tests [...] would be measured’.<sup>46</sup> Importantly, while Terman argued that the norming process had been representative, it nonetheless amounted to a test of what white middle-class children might be familiar with.<sup>47</sup>

Terman marketed the Stanford–Binet for widespread use in schools, offering teacher training programmes in test administration, and

---

43 Ibid., p. 187.

44 Chapman, *Schools as Sorters*, p. 48.

45 Ibid., pp. 22–27; Raymond E. Fancher, *The Intelligence Men: Makers of the IQ Controversy* (Norton, 1987), pp. 134–35.

46 Ibid., pp. 139–41, qt. on p. 140; Chapman, *Schools As Sorters*, p. 28.

47 Carson, *The Measure of Merit*, pp. 187–88.

publishing *The Measurement of Intelligence* (1916), a manual for teachers and other school officials.<sup>48</sup> This text evoked a range of uses for the test: from the ‘proper handling of subnormals on the one hand and gifted children on the other’, along with general educational tracking and ‘vocational guidance’.<sup>49</sup> In both *The Measurement of Intelligence* and its companion, *The Stanford Revision and Extension of the Binet–Simon Scale* (1917), Terman promoted the Stanford–Binet as an instrument that revealed inherent capacities of learners. Terman’s large set of test results yielded a Gaussian bell curve, something he mistakenly took to reflect a stable (i.e., fixed) biological quality distributed normally across a population of individuals with average, inferior, and superior “endowments”.<sup>50</sup> Such thinking followed from a history of supposition, dating at least to Francis Galton, that normal distributions indicated stable underlying natural causes of those measured differences.<sup>51</sup> The Stanford–Binet IQ test, Terman argued, could separate nature from nurture, identify a learner’s hereditary potential, and permit a more efficiently functioning school to ‘grade the raw material with which education works’.<sup>52</sup>

Terman was beginning to claim that his test could also isolate hereditary racial differences in intelligence. Discussing case studies of low measured IQ among children of Portuguese immigrants, Terman cast a broader typological net that drew together both race and vocation. Such a ‘low level of intelligence’, he observed, was also ‘very, very common among Spanish-Indian and Mexican families of the Southwest and also among negroes. Their dullness seems to be racial or at least inherent in the family stocks from which they come. High grade moron[s]’ like these could not ‘[...] master abstractions but they can often be made efficient workers’. These [were] ‘the world’s “hewers of wood and drawers of water”’.<sup>53</sup>

48 Chapman, *Schools As Sorters*, p. 62; Lewis Madison Terman, *The Measurement of Intelligence: An Explanation of and a Complete Guide for the Use of the Stanford Revision and Extension of the Binet–Simon Intelligence Scale* (Houghton Mifflin, 1916), vii, ix.

49 Ibid., vii.

50 Lewis M. Terman, *The Stanford Revision and Extension of the Binet–Simon Scale for Measuring Intelligence* (Warwick and York, Inc., 1917), pp. 26–49, 26, 39, 55.

51 Carson, *The Measure of Merit*, p. 165;

52 Terman, *The Measurement of Intelligence*, pp. 19–21, qt. on 20.

53 Ibid., pp. 90–92, qts. 91–92.

## Smith-Hughes and Vocational Education: Curricular Differentiation within the Public Schools

Terman's view of schools as separators of "raw material" would play an important role in the educational efficiency movement and the push for scientific management of schools. These efforts had also already been enjoined by a consortium of interests spanning industry, and educational administration, and were tied to calls to differentiate the curriculum by creating vocational education in the public schools. Such a programme had long been supported by major corporate capital and industrialists like J.P. Morgan and John D. Rockefeller, and trade associations like the National Association of Manufacturers.<sup>54</sup> The National Education Association (NEA), the largest national professional union for educators, was likewise receptive to such industry influence over public school curriculum.<sup>55</sup>

The political motives behind these efforts were multiple but inter-related. Industry sought more control over the training of their own labor force, which could in the process be more readily "de-skilled" for the assembly line. Getting industry-led vocational programmes into schools would prise manual training for trades free from what were then union-driven apprenticeship systems. Vocational education could then be reformulated around factory and industry-friendly principles of scientific management and Taylorism.<sup>56</sup> School officials likewise embraced vocational education reform as a way to stratify and "manage" diversity and overenrolment in schools. Vocational curriculum, its advocates argued, would serve students who were not academically inclined, offering a way of 'making more productive' poor, non-white, and immigrant students.<sup>57</sup>

In 1917, Congress responded to this menagerie of lobby interests by passing the Smith-Hughes Act, which targeted funds for formalized

---

54 Samuel Bowles and Herbert Gintis, *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*, Reprint edition (Chicago, Ill: Haymarket Books, 2011), p. 192.

55 Jeannie Oakes, *Keeping Track: How Schools Structure Inequality* (Yale University Press, 2005), p. 30.

56 Bowles and Gintis, *Schooling In Capitalist America*, p. 193. Oakes, *Keeping Track* (2005), p. 31.

57 *Ibid.*, pp. 30–32; Bowles and Gintis, *Schooling In Capitalist America*, pp. 191–92.

vocational education programmes. This move was unprecedented given the long history of resistance to federal funding of education.<sup>58</sup> Smith-Hughes was swiftly followed by the NEA's Cardinal Principles Report (1918), which argued that such vocational programmes should be established within "comprehensive" public high schools, as an alternative to separate vocational institutes. Here, students would be differentiated according to abilities, interests and likely future work. The targeted monies helped existing schools create a second vocational tier distinct from their general academic curriculum. Over the next decade, Smith-Hughes-funded curricular differentiation ushered in an era of more widespread educational tracking.<sup>59</sup>

Yet to make decisions about which students "belonged" in which curriculum, schools needed to determine their *a priori* or inherent abilities. Here, Terman's new IQ test would come to play a decisive role. There was, even by 1916, a significant and growing educational market for them.<sup>60</sup> Recall, though, that testing itself was still, for the most part, administered one test-taker at a time. Widespread adoption of IQ in schools thus awaited innovations that would allow implementation on a mass scale.

### Wartime Group Testing and National "Intelligence": The Scaling Up of IQ

Such an opportunity for the testing movement came in the spring of 1917 with the US entrance into World War One. Military ranks increased more than tenfold over the next year and a half, impelling the army to seek a means for rapidly sorting its new recruits.<sup>61</sup> Psychologist Robert Yerkes, then president of the American Psychological Association (APA), seized this chance, and gathered the leading figures in intelligence testing to design a test to meet the army's needs. The team, including

---

58 Joel Spring, *The American School: From the Puritans to the Trump Era* (Routledge, 2018), p. 257; Kaestle and Smith, 'The Federal Role in Elementary and Secondary Education, 1940–1980', p. 388.

59 Oakes, *Keeping Track* (2005), pp. 32–35; Herbert M. Kliebard, *The Struggle for the American Curriculum, 1893–1958* (Psychology Press, 2004), pp. 95–132; Bowles and Gintis, *Schooling In Capitalist America*, pp. 194–195.

60 Carson, *The Measure of Merit*, p. 178.

61 Zenderland, *Measuring Minds*, p. 281; Carson, *The Measure of Merit*, p. 228.

Goddard, Terman, and Guy Whipple, among others, met at Vineland to convene their work over the summer of 1917.<sup>62</sup> One of Terman's graduate students had already been working on a method to convert the Stanford-Binet into a paper-and-pencil, multiple choice group test that could be administered to hundreds of subjects at once.<sup>63</sup> By the end of January 1919, their new army Alpha and Beta tests (a variant for those who could not read or write) had been given to over 1.75 million new recruits representing a broad swath of US society.<sup>64</sup>

While impressive in its rapidity and scale, the published results of this wartime testing were also consequential. Testers were surprised by low scores overall. The average mental age of white enlistees was determined to be 13.08 years. Additional analysis by race and region appeared to be just as telling. Forty-seven percent of whites and eighty-nine percent of African Americans qualified as "morons", or in other words, barely above feeble-minded. Southern African Americans—living under Jim Crow segregation—scored lowest out of all groups.<sup>65</sup>

World War One army testing yielded not only a mountain of data, but also a new generation of intelligence testers. The project had required training a small army-within-the-army of young psychologists—nearly 400 strong—including Carl Brigham (creator of the SAT) and David Wechsler, among many others.<sup>66</sup> Beyond these instrumental and professional developments, intelligence—as a controlling idea—had catapulted by the war's end to the forefront of an alarmist national discourse that could be used to question the state of American civilization, and the fitness of the general population for democratic self-governance.<sup>67</sup> Numerous follow-up analyses of World War One army testing were also quick to re-assemble the mass of new data in support of older forms of race science, fuelling eugenic anxieties about low 'national intelligence', integration, immigration, 'miscegenation', and national degeneration. Carl Brigham, for example, in his widely-read *A Study of American Intelligence* (1923), drew on World War One testing to argue that descendants of Northern European nationalities

---

62 Ibid., p. 202.

63 Zenderland, *Measuring Minds*, p. 285.

64 Carson, *The Measure of Merit*, p. 218–219; Zenderland, *Measuring Minds*, p. 288.

65 Ibid., p. 289.

66 Chapman, *Schools as Sorters*, p. 74; Carson, *Measure of Merit*, p. 242.

67 Ibid., p. 232.

(England, Holland, Scotland, Germany, and Denmark)—and the so-called Nordic “racial stock” they allegedly embodied—represented the pinnacle of human intelligence. Descendants of South Eastern European races, those many recent immigrants to the US including Greeks, Russians, Italians, Poles, fell far down the list, with the “Negro draft” at the bottom rung of Brigham’s ladder.<sup>68</sup>

Brigham’s overtly race-scientific claims were no exception, but rather the norm. The early 1920s saw a slew of such attempts to re-explain the social order according to the new science of intelligence and the natural differences among peoples it alleged to uncover.<sup>69</sup> Soon enough, World War One army test data would be supplied as evidence in Congressional testimony, leading to the passage of the Immigration Restriction Act of 1924 which established sharply limited quotas for South Eastern Europeans on presumptions of racial inferiority.<sup>70</sup>

### Expanding Contexts: Selling Group Testing to US Public Schools

The post-World-War-One era saw intelligence testing likewise finally become mass practice in US schools. In 1919, Terman, Yerkes, Whipple and others secured funding from the Rockefeller Foundation to develop the National Intelligence Tests (NITs), intended, in their words, as a standardized ‘intelligence scale for the group examination of school children—a scale for the measurement of native ability’.<sup>71</sup>

The NITs were group tests developed directly from the army Alpha (itself developed from the Stanford–Binet), and designed to be administered by teachers and school officials. Though they were soon

68 Nell Irvin Painter, *The History of White People*, Reprint (W. W. Norton & Company, 2011), pp. 284–88; Gould, *The Mismeasure of Man*, pp. 254–60; Carl Campbell Brigham, *A Study of American Intelligence* (Princeton: University Press, 1923), pp. 146 169, 170.

69 Carson, *The Measure of Merit*, pp. 242–43. See for example Lothrop Stoddard, *The Revolt against Civilization: The Menace of the under Man* (C. Scribner’s Sons, 1922); William McDougall, *Is America Safe for Democracy?: Six Lectures Given at the Lowell Institute of Boston, Under the Title “Anthropology and History, Or The Influence of Anthropologic Constitution on the Destinies of Nations,”* (C. Scribner’s Sons, 1921), see for example pp. 51–58, 63–64.

70 Graham Richards, *Race, Racism and Psychology: Towards a Reflexive History* (Routledge, 2012), pp. 90–92, 105; Gould, *The Mismeasure of Man*, pp. 261–262.

71 Chapman, *Schools as Sorters*, p. 77.



standardized for third grade through high school, their norms were generated using a largely white, middle-class, non-rural cross-section of American society.<sup>72</sup> Despite this biased test construction process, Terman and his co-authors wasted little time pitching their new group test to publishers. It would cut the ‘time and labor’ of testing to ‘one twentieth of the present requirement and would be demanded by the million’, they argued. The tests were soon bought by World Book Publishers and turned in a very successful initial run, with some 400,000 tests already in circulation by the fall of 1920.<sup>73</sup>

The advent of group testing—in the wake of Smith-Hughes-funded curricular stratification—created an educational mass market for intelligence tests, which many psychologists were eager to exploit. The NITs were, of course, only one among many new proprietary group tests patterned after the Stanford–Binet.<sup>74</sup> US psychology itself began to organize around this lucrative new private enterprise. In 1921, Terman joined forces with psychologist James McKeen Cattell to form the Psychological Corporation, with offices opening in many states, to supply a range of applied psychological instruments and testing services. By 1929, half of the Corporation’s income came from marketing intelligence tests.<sup>75</sup> Psychologists directly appealed not only to publishers but to educational officials and policy makers as well, selling the policy relevance of intelligence testing along with the IQ hereditarianism that these tests had come to embody.

What about the growth of school intelligence testing in actual practice? Chapman provides clear evidence of systematic adoption of testing, grouping, and tracking across the 1920s in the three California Bay Area school systems that formed the basis of his case study. He also demonstrates that these practices were commonly accompanied by assertions from school officials that such intelligence tests offered ‘scientific proof of [the “racial”] inferiority’ of non-whites, and that preparation for different kinds of labor (“brain work” or “hand work”) required differentiation on the basis of intelligence. There seems little

72 Ibid., pp. 77–81, 148.

73 Ibid., pp. 81–82, qt. on 81.

74 Ibid., pp. 92–101, 98.

75 Ibid., p. 105; Michael M. Sokal, ‘The Origins of the Psychological Corporation’, *Journal of the History of the Behavioral Sciences*, 17.1 (1981), 54–67 (pp. 54, 61), [https://doi.org/10.1002/1520-6696\(198101\)17:1<54::AID-JHBS2300170108>3.0.CO;2-R](https://doi.org/10.1002/1520-6696(198101)17:1<54::AID-JHBS2300170108>3.0.CO;2-R)

doubt that an educational management system had emerged with the potential to disproportionately direct immigrant, non-white, and lower-class students toward Smith-Hughes funded vocational education curricula—all under the “objective” criteria of measured intelligence.<sup>76</sup> By the early 1920s, the testing movement had already become quite diffuse and was promoted by not only Terman, Goddard, Yerkes and the like, but now also by thousands of educational and psychological professionals nationwide.<sup>77</sup> This broader, if still decentralized, embrace of psychometric entrepreneurialism by “efficiency-minded” school administrators was increasingly sustained and monetized by a growing testing market and through the incremental habituation of new school practices.

In 1925, the Bureau of Education conducted a national study on testing and grouping in schools in cities with a population greater than 10,000. Their survey indicated that in responding school systems, ‘nearly two-thirds of the elementary schools, over one-half of the junior high schools and two-fifths of the high schools’ used intelligence tests to group or track students according to “ability”. The diagnosis of so-called “subnormal” or feeble-minded individuals likewise remained a highly prioritized use of tests.<sup>78</sup> By 1922, Terman had been elected president of the APA and maintained a great deal of sway over the testing movement through the 1920s and 1930s. Terman’s many successes in test publishing had likewise made him quite wealthy.<sup>79</sup>

## Testing in Interwar Norway by Comparison

Broadly, the growth of school IQ testing in Norway was altogether slower than in the US, and the eventual scope of the uses of IQ were more limited. As in the US, the origins of Norwegian intelligence testing lay in marking out the domain of “sub-normality.” When testing for special education diagnosis began in the early 1900s, it was sparse,

76 Chapman, *Schools as Sorters*, pp. 107–27, qts. on 118, 124.

77 Ibid., p. 146. See for example Stephen Petrino, “The “Never-to-Be-Forgotten Investigation”: Luella W. Cole, Sidney L. Pressey, and Mental Surveying in Indiana, 1917–1921”, *History of Psychology*, 4.3 (2001), 245–71, <https://doi.org/10.1037/1093-4510.4.3.245>

78 Chapman, pp. 155–59. This survey had a thirty-five percent response rate.

79 Ibid., p. 105; Fancher, *The Intelligence Men*, p. 141.

localized, and conducted with translated (not standardized) versions of individual tests like the Binet–Simon. It was only in the 1920s that serious attempts were made to standardize individual IQ tests for a Norwegian population. Similar work with group tests began a decade later. These notable developments followed in the wake of somewhat protracted educational reform efforts, begun in 1889 and fulfilled in the 1936 Folk School Act, with a common seven-year national curriculum in 1939.<sup>80</sup> Legislation during this period diverted state funds away from more elite educational options available to the upper-middle classes, and concentrated public resources in schools offering undifferentiated, common curricula for ages seven to fourteen.<sup>81</sup>

Yet, as noted, this effort to modernize and expand schooling for all compelled boundary-making around the “normal” general school population, and a separating out of those students deemed “weakly gifted”. Thus, as Ageliki Lefkaditou (chapter 1, this volume) demonstrates, early calls for the seven-year reform mark an acceleration in test development efforts. The first IQ test normed for a Norwegian population was a 1931 standardization of Terman’s Stanford–Binet. This first was eventually followed in 1954 by a standardization of Terman’s updated Stanford–Binet (1937).<sup>82</sup> These 1931 and ‘54 standardizations mark important thresholds in the Norwegian expansion of individual IQ testing for special education—a practice that could be said to have finally been established as systematic and national with the ‘54 standardization and the co-emergence of the Educational Psychological Counselling Services.<sup>83</sup>

Such was the comparatively slow expansion of individual IQ testing in Norway. The story of interwar Norwegian group testing is somewhat more nuanced, equally unhurried, and ultimately more tenuous. Even so, the eventual outcomes of these forays with class-administered pencil-and-paper tests are not at all insignificant. Scattered group testing began

80 A decisive step in this process was the 1920 parliamentary resolution to limit state funding to the seven-year common curriculum, while municipalities or parents’ associations could still finance other forms of schooling.

81 Larsen, “‘Med måling følger viden’ Intelligensmåling i den norske folkeskolen 1900–1945”, pp. 45–46; Val Dean Rust, *The Democratic Tradition and the Evolution of Schooling in Norway*, Contributions to the Study of Education (Greenwood Press, 1989), pp. 103–190.

82 See Paulsen, chapter 9, this volume.

83 See Ludvigsen, chapter 6, and also Paulsen, chapter 9, this volume.

in a handful of schools around the country as early as the 1920s, but such instances involved unstandardized, heirloom translations produced by and exchanged among individual psychiatrists and pedagogues.<sup>84</sup> Moreover, unlike mass testing already underway in the US, such very early Norwegian group testing was rarely, if ever, intended for stratifying students within general educational settings. Rather, these tests were auditioned here for a range of purposes, including assessment of the tests' own validity and predictive value in relation to other school exams—and occasionally as a measure of “maturity” or school readiness in very young children.<sup>85</sup>

Yet, as already indicated, there are important exceptions to this account of interwar Norwegian group testing as sporadic and exploratory. From 1934 into the 1950s, Norwegian standardizations of US group tests were used in Oslo and in some other large municipalities as a screening device for special education.<sup>86</sup> The reasons for these exceptions possibly lie in the size and the diversity of the student-aged population in Norway's largest school systems, along with matching organizational resources there to carry out such tests. Students identified by the group test as candidates for institutionalization or ‘helping classes’ would then be re-tested individually with the Norwegian standardization of the Stanford–Binet.<sup>87</sup> More than this, there is clear indication of an emergent practice of group testing for general ability differences among so-called normal students, again in some larger Norwegian towns and cities, most prominently Oslo. Though still unsystematic compared to the US case, such instances of general ability testing likely led to some individualized instructional modifications, made at the discretion of teachers or schools. It remains unclear, however, the extent to which any such pedagogical practices followed from test results.<sup>88</sup> It is also worth noting that the ages of Norwegian students who might have been exposed to general ability testing largely overlapped with their US counterparts, though they may not have precisely aligned in all cases. The average age of minimum mandatory schooling in the US was sixteen, while in Norway, until 1969, it was fourteen.

---

84 Asbjørn Birkemo, *Pedagogisk-psykologisk rådgivning* (Oslo: Universitetet i Oslo, 2009).

85 Ibid.

86 Larsen, ‘Med måling følger viden’, pp. 50–51.

87 Birkemo, *Pedagogisk-psykologisk rådgivning*, p. 37.

88 See, for example, Thue, chapter 11, this volume.

As in the US, the growth of Norwegian testing—both group and individual—was certainly energized by policy. Yet, while Norway’s interwar educational reforms forged a demarcation between special and general education, they differed decidedly from their US counterparts (namely Smith-Hughes and Cardinal Principles), by upholding a common curriculum and not enforcing differentiation across the general student population. Thus IQ testing for general ability in interwar Norway had no bearing on who pursued what kind of schooling or for how long. Isak Emberland’s analysis of the role of testing in Norway’s stand-alone Vocational Schools (chapter 5, this volume) heightens this point of contrast. Here, IQ was not used to track or involuntarily assign Norwegian students in the common Folk School to a vocational curriculum. Moreover, while American titans of industry bent US vocational education policy to their ends, it was Norwegian labor—guilds and trade unions—that maintained shaping power over Norwegian Vocational Schools and their curriculum.<sup>89</sup> The top-heavy influence of capital and market actors over educational policy in the US case is absent in the Norwegian history.

Relatedly, Norwegian test development itself was not fuelled by major state or corporate-philanthropic funding. Significantly, on this score, there was no Norwegian state- or military-funded wartime mobilization of test development during the World War One period.<sup>90</sup> Moreover, as Gard Paulsen observes (chapter 9, this volume), there was likewise little to no proprietary test publishing market in interwar (or, for that matter, post-World-War-Two) Norway, which might otherwise have accelerated such efforts. Norwegian test development was thus decidedly “bottom up”, relying on the commitments and resources of interested individuals and associations across pedagogy and psychiatry.

<sup>91</sup> Indeed, such efforts sometimes drew from the pockets of individual

---

<sup>89</sup> See Emberland, chapter 5, this volume.

<sup>90</sup> Psychologist Martin L. Reymert worked on a Norwegian standardization of the army Alpha test after World War One, but never finished it. See Lefkaditou, chapter 1, this volume.

<sup>91</sup> Ydesen, Ludvigsen, and Lundahl, ‘Creating an Educational Testing Profession in Norway, Sweden and Denmark, 1910–1960’, p. 125; Larsen, ‘Med måling følger viden’, p. 66. This, however, does not imply that the Norwegian state was absent, as several people involved in test development were public servants, and their work was supported by state institutions and scholarships. On this see Lefkaditou, chapter 1, this volume.

Norwegian testers themselves.<sup>92</sup> It is this comparative lack of synergistic interaction between policy and market that appears, at least in part, to explain the slower pace and more limited scope of testing in Norway relative to the US.

Nonetheless, the efforts of these early testers clearly had important downstream effects on later practice. Norwegian psychiatrists and pedagogues were, by the mid-30s, gradually securing more institutional support for their ongoing research with intelligence.<sup>93</sup> Individual IQ testing for special educational diagnosis was also, by then, well on its way to becoming established national practice. More than this, the numerous—yet comparatively unsystematic—experimentations with group testing began to play an important role in shaping expert attitudes about ability or intelligence across the general population. Such episodes of general IQ sampling had, in fact, begun to convince many Norwegian testers that there was a wide range of ability (effectively understood as degrees of natural capacity) among the “normal” student-aged population. Researchers poring over results of Oslo group testing, for example, entertained but ultimately dismissed the notion that the IQ variance they observed might be explained by class bias or socioeconomic disparities. Instead, the tests appeared to them to provide objective measures of widespread natural inequality.<sup>94</sup> Such emergent perceptions of population-level hereditary individual difference would later support mounting calls for curricular differentiation in the post-World-War-Two era.<sup>95</sup>

## Criticisms and Accommodations in the US

The American story has thus far surveyed the brisk march of intelligence across the scientific, cultural, and educational terrain of the early twentieth-century US. This new practice certainly did not go uncontested, however, and indeed numerous criticisms—particularly of its hereditarian and racio-biologic claims—were levelled against it. Yet, as the final portions of this chapter will suggest, such concessions as

---

92 Paulsen, chapter 9, this volume.

93 Helsvig, chapter 12, this volume. Larsen, ‘Med måling følger viden’, p. 68.

94 *Ibid.*, pp. 50–51, 77.

95 *Ibid.*, p. 77. Also Helsvig, chapter 12, this volume.

these scientific disputes managed to secure were in the end limited or flexible enough for the US testing movement to sooner or later pivot around and move beyond.

An early challenge to testing came from psychologist J. E. Wallace Wallin who, at a 1915 APA gathering, argued that intelligence testing had unfairly targeted and misunderstood institutionalized and incarcerated populations. Wallin had screened successful adults from across a wide swath of the American public and found that a surprisingly high rate of these well-adjusted individuals scored as feeble-minded. For Wallin, this suggested that feeble-mindedness was a category propped up by dubious assumptions—about intelligence, about what test scores indicated—and then applied to populations who lacked the social resources to dispute the judgments levelled against them.<sup>96</sup> Such inconsistencies also suggested to many that intelligence testing was likely saddled with an unresolvable definition problem. Just what was being measured and what was being claimed on the basis of these measurements? None of the mainstream testers had arrived at anything like a scientifically operationalizable definition of what intelligence actually was. How could one pretend to quantify something seemingly so all-determining and yet not bother to clearly stake out the dimensions of this object of measurement?<sup>97</sup>

There was moreover the observation that, absent a scientific definition, IQ was merely an index of normative competencies. Test content was, in other words, both selective and arbitrary, implicitly reproducing not intelligence but testers' own more narrow assumptions about what constituted valuable knowledge. A range of psychologists across the early teens thus noted that tests could be significantly biased by social factors such as class, ethnicity, educational exposure, and linguistic differences.<sup>98</sup> Binet himself had noted that such biases amounted to an 'intrusion' of social, contextual and educational factors that deeply confounded the separability of "nature" from "nurture" in regard to measured differences. One could not know how much individual IQ

---

96 Zenderland, *Measuring Minds*, p. 246; J. E. Wallace Wallin, 'Who Is Feeble-Minded?', *Journal of the American Institute of Criminal Law and Criminology* 6.5 (1916), 706–716 (pp. 710–11), <https://doi.org/10.2307/1133344>.

97 Zenderland, *Measuring Minds*, p. 129, 159; Carson, *The Measure of Merit*, p. 186, 258.

98 *Ibid.*, pp. 179, 186; Zenderland, *Measuring Minds*, p. 241.

variance was due to putative natural differences.<sup>99</sup>

Binet's agnosticism seems to have been more or less lost (with a few exceptions) on US psychologists and educators alike.<sup>100</sup> Most of the mainstream American testing community expressed increasing confidence that tests could isolate innate "capacity". IQ's critics, however, were quick to observe that such claims could be arrived at only after studiously ignoring or inadequately ruling out social, educational, or other environmental factors.<sup>101</sup>

Public intellectual Walter Lippmann (also agnostic on the matter of nature and nurture) pointed out that Terman's carefully burnished bell curves—and the IQ-stability and validity he attributed to them—were a statistical artefact, not the discovery of an underlying population-wide natural distribution of mental ability.<sup>102</sup> Later critics have elaborated that the routine test construction process itself required that tests reproduce a Gaussian distribution.<sup>103</sup> Questions were retained if they helped produce a normal curve, and discarded if they did not. The IQ curve was made, in other words, not born. Lippmann also advanced that there appeared to be a link between the heritability claims of testers and the marketing of tests. 'The intelligence test', he observed, 'is being sold to the public on the basis of the claim that it is a device that will measure pure intelligence'.<sup>104</sup> Tests were marketable (and 'policy-relevant') in a US context to the extent that they could be said to isolate natural differences. This was the pretext schools required for sorting.

This question of the market for testing (and also its regulation) amounted to yet another arena of concern. Guy Whipple and Wallin, for example, were by 1915 already worried about the rapidly growing

99 Ibid., p. 129, 241; see also Evelyn Fox Keller, *The Mirage of a Space between Nature and Nurture* (Duke University Press, 2010).

100 Ibid., p. 130.

101 Walter Lippmann, 'Tests of Hereditary Intelligence', *New Republic*, 32.416 (22 November 1922), 328–330 (pp. 329–330).

102 Walter Lippmann, 'The Reliability of Intelligence Tests', *New Republic*, 32.414 (8 November 1922), 275–277 (p. 275); Terman, *The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence*, p. 42.

103 Lynn Fendler and Irfan Muzaffar, 'The History of the Bell Curve: Sorting and the Idea of Normal', *Educational Theory*, 58.1 (2008), 63–82 (p. 63), <https://doi.org/10.1111/j.1741-5446.2007.0276.x>; Michael J. A. Howe, *IQ in Question: The Truth about Intelligence* (SAGE, 1997), p. 20; see also Lefkaditou, chapter 1, this volume.

104 Walter Lippmann, 'A Future for the Tests', *New Republic*, 33.417 (29 November 1922), 9–10 (p. 9).



number of “amateurs” (i.e., teachers, school administrators, and other non-psychologists) among practicing testers. Yet advocates like Terman refused to rein in the training of new teacher-testers, even though the APA would soon adopt a formal referendum against such practice. More teacher-testers, of course, meant more testing overall, thus more test sales and royalties. Whipple was soon forced to acknowledge that public demand and market forces—not professional guidelines—would set the future course of testing.<sup>105</sup>

The widespread assertions of white (and also specifically Nordic) intellectual superiority following testing in World War One also generated, across the mid- to late 1920s, focused criticism of these claims of differential racial IQ. Numerous social scientists including psychologist Otto Klineberg, African American sociologist Horace Mann Bond, and anthropologists Franz Boas and Margaret Mead demonstrated through their own counter-analyses and experiments the clear effects of education and social environment—including racism itself—on measured IQ within and across different race groups. This cohort of anti-racist scientists established that test content generally comprised white, middle-class knowledge norms, and that the large degree of overlap between white and non-white IQ distributions itself strongly suggested that there was no categorical (i.e., biological) difference among the so-called races in relation to intelligence.<sup>106</sup> By the mid-1930s, it was widely accepted among psychologists and social scientists that claims of differential racial IQ (and white intellectual superiority) had no basis in science. By 1930, Carl Brigham had disavowed his prior assertions of race difference. Terman and Yerkes quietly suspended theirs.<sup>107</sup>

The growing attention paid to the role of the environment had begun to motivate those at the leading edge of scientific debate to challenge bedrock claims of the heritability and stability of IQ at the level of the individual. By the mid-1930s, a group of psychologists at the Iowa Child Welfare Research Center had produced a coherent body of research showing that measured IQ varied a great deal—in many cases improving

105 Zenderland, *Measuring Minds*, p. 249.

106 Richards, *Race, Racism and Psychology*, p. 127; Carl N. Degler, *In Search of Human Nature: The Decline and Revival of Darwinism in American Social Thought* (New York: Oxford University Press, 1992), pp. 67, 84–104, 176, 181.

107 Carson, *The Measure of Merit*, p. 261; Richards, *Race, Racism and Psychology*, p. 104; Gould, *The Mismeasure of Man*, pp. 221–222.

from feeble-minded to normal—when individuals were followed across changing social circumstances: from orphanage to adoptive home and stable school setting, for example. They concluded that individual IQ was not a fixed, inborn capacity, but subject to change, depending on the environment.<sup>108</sup> A clash with Terman and the mainstream psychometric community over these findings appeared inevitable.

Observers of the ensuing and contentious debate were eager to call a draw, and from 1940 onward, the whole field adopted the conciliatory language of “nature-and-nurture” interactionism. But just how did either set of causes contribute, and how much variance in measured difference was due to nature or to nurture in any individual case? No one knew, but now the role of nurture (and the countless social, educational, and cultural factors that nurture might subsume) could not be discounted outright. Yet, despite the new interactionist language, Terman and mainstream psychometrics remained inclined toward hereditarian explanations for measured individual difference.<sup>109</sup> Hereditarian IQ—the belief that made testing so marketable as a sorting device—lumbered on.

Even so, the conceptual and linguistic compromises produced by these “race IQ” and nature-nurture debates help explain the state of testing and expert language about intelligence (or ability) in ongoing policy making contexts following World War Two. Firstly, now, at least in expert discourse, the locus of natural difference could be said to lie only at the level of the individual, not the race group. Secondly, the more intensive focus on individual IQ that ensued was now tempered with apologetics for a seemingly fair-minded nature-nurture interactionism. With such cautions now in place, IQ’s champions argued testing would be “color-blind” and fairly identify the “bright” from across all segments of American society, even from disadvantaged groups that might otherwise have been overlooked.

But *none* of the fundamental problems with testing had actually been resolved. The problem of defining “intelligence” persisted, as did the inability of IQ to isolate “nature”. Tests remained decidedly biased

---

108 Hamilton Cravens, *Before Head Start: The Iowa Station & America's Children* (Chapel Hill: University of North Carolina Press, 1993), pp. 132–34, 173–74, 182–183, 212–214.

109 *Ibid.*, p. 215.

by race and class in favor of middle- and upper-class white people.<sup>110</sup> More than this, new avowals of nature-nurture interactionism could easily default to hereditarianism in both conception and practice. This slippage is evident even in the generic argument for color-blind testing made above. Here, the matter of differential “brightness” appears as established fact. Brightness as potential or capacity is assumed to be already formed, inherent in select individuals *a priori*, and not the result of ongoing interaction or education. A brief consideration of testing policy in the decades after World War Two is revealing in this regard.

### Intelligence post-World War Two: *Brown*, the NDEA, and “Color-Blind” Hereditarian Individualism

In the immediate wake of the Soviet Union’s 1957 Sputnik launches, the US Congress passed the National Defense Education Act (NDEA, 1958) which authorized federal funds for, among other things, greatly expanded and more systematic ability testing in high schools across the country. The NDEA allocated finances to schools to purchase tests, start or resume testing programmes, and hire guidance counselors to help enforce test-based placement decisions. The NDEA also designated monies for the strengthening of high school science and math instruction, and was accompanied by expert recommendations urging high schools to limit rigorous science and math classes to the top fifteen percent of students as determined by measured ability.<sup>111</sup> These NDEA programs were directly followed by the creation of a new college preparatory high school science and math curriculum funded by the National Science Foundation (NSF).<sup>112</sup>

Here, then, was a comprehensive suite of initiatives that offered schools resources to improve their curricular on-ramp to college, and

110 Keith M. McNamara, ‘Testing Bias in Psychology, Law, and Public Policy, 1920–1980’ (PhD Dissertation, United States—Wisconsin, University of Wisconsin-Madison, 2022), pp. 6, 160–168, 174–241, 334–344.

111 Barbara Barksdale Clowse, *Brainpower for the Cold War: The Sputnik Crisis and National Defense Education Act of 1958* (Greenwood Press, 1981); James Bryant Conant, *The American High School Today: A First Report to Interested Citizens* (McGraw-Hill Book Company, 1959).

112 John L. Rudolph, *Scientists in the Classroom: The Cold War Reconstruction of American Science Education* (Palgrave Macmillan, 2002).

also to regulate who got access to it. This NDEA-NSF policy complex has been widely understood as a Cold War reform effort, catalyzed by the Sputnik crisis, to improve the nation's long-term scientific and technological competitiveness by educating its best and brightest in these fields.

But there were more than just Cold War forces at play in these developments. Planning and drafting for the NDEA actually began in 1955, well before the Sputniks, and in fact just months after the 1954 *Brown v. Board Supreme Court* decision impelling racial integration of US public school systems.<sup>113</sup> *Brown v. Board* applied most directly to Southern schools still under *de jure* segregation, but it also implicated many other school systems across the country which maintained segregation either by flouting state laws or by adjusting school district zones to established patterns of residential segregation. The *Brown* decision was followed by widespread white backlash (north and south), including violence, threat of violence, organized "Massive Resistance" campaigns, and opposition to integrative busing. The conflict would roil the nation's system of education, and NDEA-era policy makers were eager to find ways to stem a potential tide of "white flight" from public schools.<sup>114</sup>

The effects of the NDEA and *Brown* should be considered together, given their overlapping chronologies and contexts. As a result of biases in testing by ethnicity and class, and stark generations-long educational inequalities produced by racially segregated schooling, NDEA-funded ability testing and placement programmes could recreate educational divisions by race and class within schools, just as newly integrating African American students began to enter what had been "white schools".<sup>115</sup>

113 Wayne J. Urban, *More Than Science and Sputnik: The National Defense Education Act of 1958* (University of Alabama Press, 2010).

114 Davison Douglas, *Jim Crow Moves North: The Battle over Northern School Segregation, 1865–1954* (New York, NY: Cambridge University Press, 2005); John Kyle Day, *The Southern Manifesto: Massive Resistance and the Fight to Preserve Segregation* (Univ. Press of Mississippi, 2014); John Gardner, 'Observations on the Conant Study (Keppel's and Others' Feedback, Compiled and Edited by John Gardner)', n.d. Papers of James Bryant Conant, 1862–1987, UAI 15.898, High School March - June 1957, Box 42. Harvard University Archives.

115 Jim Wynter Porter, 'A "Precious Minority": Constructing the "Gifted" and "Academically Talented" Student in the Era of Brown v. Board of Education and the National Defense Education Act', *Isis*, 108.3 (1 September 2017), 581–605, <https://doi.org/10.1086/694446>

Yet policy-makers who promoted the NDEA strictly observed the race-neutral language of individualized intelligence (or ability). Testing, they argued, only identified differences at the level of individuals, and because it was objective, bright individuals might be discovered from any background: any race, culture, creed, or socio-economic group. Such experts likewise typically made overtures to nature-nurture interactionism and avoided the overtly hereditarian interwar language of “inherent” or “natural” intelligence. Some made arguments for sorting instead on the basis of developed ability, a locution that could neatly accommodate both nature and nurture.<sup>116</sup> Others adduced bell curve reasoning or similar statistical argumentation to convey that academically talented individuals occurred at a fixed and dependable frequency as part of a national distribution of ability.<sup>117</sup> “Academic talent” or “intelligence,” this line of argument suggested, was already out there, pre-formed, like a natural (and national) resource. But these were calls for ability-based sorting and grouping all the same, and ability—whether developed or inherent—was still held to be more or less fixed. Education, it was implied, could not play much role in altering a student’s trajectory. Here then, in the language of post-World-War-Two NDEA policy-making, was a race-neutral individualism, garbed in the language of nature-nurture interactionism, that nonetheless was still hereditarian in practice and conception, and which could still discriminate by race and class.<sup>118</sup>

These policies emerged against the demographic backdrop of the post-World-War-Two baby boom and looming over-enrolments. US public schools remained predominantly locally funded, and in general,

116 Henry Chauncey, ‘Hearings before the Committee on Education and Labor, US House of Representatives 85th Congress, 2nd Sess. on Scholarship and Loan Program’, § Committee on Education and Labor (26 February 1958), 1111.

117 Frederick Hovde, ‘Hearings before the Committee on Labor and Public Welfare, U.S. Senate, 85th Congress, 2nd sess. on Science and Education for National Defense’ (28 January 1958) 94–95; Conant, *The American High School Today*, p. 78; James Bryant Conant, Papers of James Bryant Conant, 1862–1987, ‘Correspondence with William Alexander’, 22 September, 198, UAI 15.898, Ability Grouping, Departmentalization, Box 108, Harvard University Archives.

118 Jim Wynter Porter, ‘Guidance Counseling in the Mid-Twentieth Century United States: Measurement, Grouping, and the Making of the Intelligent Self’, *History of Science*, 58.2 (1 July 2020), pp. 191–215, <https://doi.org/10.1177/0073275319874977>; Jim Wynter Porter, ‘The Entanglement of Racism and Individualism: The U.S. National Defense Education Act of 1958 and the Individualization of “Intelligence” and Educational Policy’, *Multiculturalism*, 38 (December 2018), 3–17.

under-resourced. While public schools got no new federal dollars for much-needed school construction or teacher salaries, they were now—thanks to the NDEA—flush with funds ear-marked for standardized tests of ability. The Educational Testing Service, one among many prominent test publishers, recorded skyrocketing annual returns in test sales and an elevenfold growth in net worth in the decade and a half following the NDEA.<sup>119</sup> Likewise, as subsequent educational sociology has demonstrated, the decades of expanded testing following the NDEA saw the reproduction of significant disparities in educational opportunity by race and class.<sup>120</sup>

## Testing in Norway after World War Two

What happened in Norwegian test development and practice in the years after World War Two? Modernization efforts gave rise to another round of comprehensive reforms in both special and general education, which again spurred developments in testing.

The 1951 Special Schools Act cleared the way for extensive growth in special education. A new state Directorate for Special Schools was tasked with the differentiation of disabled children and with defining their educational needs. In 1955, an amendment to the *Folkeskole* act made municipal provision of special education mandatory, and introduced state co-funding of posts for school psychologists to assess ‘school maturity, aptitude, behavioral difficulties, and academic difficulties’ of students.<sup>121</sup> 1954 saw the launching of the new Norwegian standardization of the updated Stanford–Binet, which became a key instrument for the newly established and rapidly growing Educational Psychological Counseling Services.

Calls for more IQ testing of the general student population were growing as well. Those interwar perceptions among Norwegian testers that group testing had revealed a range of natural ability among normal

---

119 James Wynter Porter, ‘Constructing the “Gifted” and “Academically Talented” Student: “Intelligence”, Intelligence Testing, and Educational Opportunity in the Era of Brown v. Board and the National Defense Education Act’ (Ph.D Dissertation, Michigan, Michigan State University, 2017).

120 Karolyn Tyson, *Integration Interrupted: Tracking, Black Students, and Acting White after Brown* (New York: Oxford University Press, 2011); Oakes, *Keeping Track* (2005).

121 See Ludvigsen, chapter 6, this volume; Caspersen, chapter 7, this volume.

students carried forward into the post-World-War-Two era. Transnational sympathies fed this momentum. Many progressive pedagogues—like Marta Sørensen, Helga Eng, and Johannes Sandven—were influenced by US and Swedish advocates of testing across the interwar and early post-World-War-Two era.<sup>122</sup> Sandven, who followed Eng as chair of Pedagogical Research at the University of Oslo had written in 1943, for instance, that intelligence testing was still ‘far too little used for measuring normal children’. He argued that systematic group testing could provide teachers with objective information about student capacities, and support ‘possible differentiation’ within general school settings.<sup>123</sup> Such aspirations awaited the educational modernization efforts of the 1950s.

In addition to legislation for special education, the early 1950s saw the emergence of plans for a new nine-year curriculum that would ensure everyone had schooling until at least the age of sixteen. These efforts were spurred by realizations that the seven-year reform—while certainly a step in the right direction—had exposed widespread inequality by class and according to region. Most students in well-resourced urban systems could continue their academic education through high school and beyond, if they desired. Yet students in rural schools most often went to work after their seven years of common schooling, or after one or two additional years at the practically oriented *Framhaldsskolen* [the Continuation School]. This was simply because other options were not readily available. Labor and Progressives saw the extended nine-year plan as a way to address this socio-economic and regional disparity.<sup>124</sup> Yet attempts to resolve this inequality at the level of class occasioned the counter-mobilization of notions of inherent difference at the level of the individual. Perhaps not everyone was suited for an extended academic curriculum; IQ and ability tests, Sandven and others argued, would help sort the general population within the new nine-year curriculum.

An episode of more systematic testing of the general school population was thus inaugurated in 1954, along the lines of Sandven’s proposals. As Kim Helsvig observes, a Parliamentary report of that year

122 Birkemo, *Pedagogisk-psykologisk rådgivning*. See also Helsvig, chapter 12, this volume.

123 Johannes Sandven, *Intelligensmåling Ved Gruppeprøver* (Oslo, 1943), p. 81, quotes trans. Jon Kyllingstad.

124 See Helsvig, chapter 12, this volume.

initiated a period of experimentation across the general school-aged population to establish a working model for how the eventual nine-year curriculum might be implemented. This educational trial ran until 1960 and involved nearly twenty schools in comprehensive programmes of ability testing and placement within a differentiated curriculum. But this effort marks the zenith of general group testing in post-World-War-Two Norway.<sup>125</sup> As Helsvig and Fredrik Thue explore in their chapters in this volume, by 1960, criticisms of mass testing had emerged from the left, right, and center of Norwegian politics, and educational differentiation by measured ability was rapidly losing support. The nine-year plan was formalized in 1969 as, once again, an undifferentiated, common curriculum. Group testing of the general population, no longer policy-relevant, had atrophied.<sup>126</sup>

By 1960, Norway's short-lived experiment with mass testing was already drawing to a close. In sharp contrast, the US, at the same moment, was well underway (post-*Brown* and post-NDEA) with its second round of expansion, this time with a very large allocation of federal funds expressly for testing and placement services themselves. The Norwegian public—including, with exceptions, school officials, experts and politicians—clearly appears not to have bought into mass testing to the extent, or with the sustained intensity, of their US counterparts.

## Comparisons and Conclusions

From the American standpoint, the doubling in US educational policy formation that echoes across the first and second half of the twentieth century is striking. During both the 1910s and the 1950s, events that would increase general ethnic and socio-economic diversity in schools were quickly followed with policy that stratified the curriculum and tracked students toward different vocational end points. Smith-Hughes compartmentalized vocational education within public schools, while the NDEA talent hunt, together with the NSF curriculum projects, helped schools both to develop and selectively regulate their college preparatory curricula. Both policy approaches functioned by race and class, with the effect of disproportionately (if not exclusively)

---

<sup>125</sup> Ibid.

<sup>126</sup> Ibid., and Thue, chapter 11, this volume.



conserving academic and pre-professional domains of the high school curriculum for middle- and upper-class white people. Intelligence and its measurement were central to policy implementation in each case, and both the Smith-Hughes Act and the NDEA were a boon to the testing market.

The trio of factors identified at the outset—policy, markets, and racism/classism—all worked together across time in this account. Racism and classism, sharpened by fiscal conservatism, shaped beliefs about what intelligence was, who was capable of what kinds of work, and contributed to anxieties about educational resource-sharing among white people. Such conditions helped beckon forth policy that promoted general differentiation and that meshed well with established economic structures which were long dependent on the exploitation of non-white and lower-class labor. These policies would play an important role in reproducing labor market conditions favorable to industry, and moreover, stimulate the growth of a testing market that would further incentivize and habituate the practice of intelligence testing. Notably, in the case of Smith-Hughes, it is clear that industry lobbies (i.e., market actors) helped broker federal policy that served their own interests. All this is indicative of the perpetuation of a form of racial capitalism—evolved from a history of slavery—that could readily co-produce race and class, and which was rendered less visible as such in schools as objectivist and allegedly race-neutral practices of intelligence or ability testing.

Interaction among these factors was sustained in significant measure by IQ hereditarianism, and the authority of experts who were its champions. While overtly hereditarian discourse was discouraged by the 1950s, conceptions of fixity had become embedded in practice and were now screened behind a language of nature-and-nurture interactionism. Yet, these so-called interactionist arguments simply found new ways to frame measured difference as an individual trait or essence that was largely fixed and predictive of later outcomes.

The trio of factors—policy, market and racism/classism—played a significant role in the US in driving the expansion of IQ testing outward from its original domain of special education diagnosis to also include sorting across the general school-age population. The other chapters in this anthology all collectively demonstrate that most of these elements were present to some extent in Norway. Yet the degree to which they

were coordinated in the maintenance of a labor regime appears to have been significantly less. Notably, on this score, there is no history in Norway of slavery, an institution that presumed to own its captive labor, which bound race and work so inhumanely together, and out of which later ethnic, class, labor, and economic relations might evolve. Where American educational policy worked to stratify and separate at all levels, Norwegian educational reforms (while entertaining a relatively brief flirtation with tracking) maintained and extended a common curriculum—at least within the boundaries of general education. Furthermore, the Norwegian policy-making process appears to have been entirely more insulated from market actors with little indication of the kind of market-policy corroboration and co-escalation apparent in the US case. There is no evidence, for instance, of interwar Norwegian industrialists bending educational policy to their own aims. Likewise, there was no comparable Norwegian test publishing market which might stoke and normalize general demand for testing, and by which testers and publishers could enrich themselves.

It appears that for these reasons—the absence of a racial project that could be specifically described as racial capitalism, along with the lack of US-style market-policy interaction—mass IQ testing of the general school-age population was effectively resisted in Norway during the late 1950s and 1960s through broad-consensus democratic politics.<sup>127</sup> IQ-driven differentiation in Norwegian education would remain largely constrained to the border between special and general education.

Despite these clear US-Norwegian differences, there remain some important continuities to sift through before this analysis rests: namely, methodological issues concerning racism, classism, and hereditarianism specifically related to IQ test construction and interpretation. There is ample evidence, for example, that some, if not many, prominent Norwegian testers subscribed to IQ hereditarianism in some form or other. Sometimes such conceptions were explicitly held, as in psychiatrist Johan Lofthus' adoption of Cyril Burt's definition of intelligence as 'inborn' and generally unresponsive to environment or education.<sup>128</sup> Yet, as in the US, adapted or implicit forms of IQ hereditarianism were also common in Norway. Sandven believed that the 'fundamental maturity

---

127 See chapters by Thue, chapter 11, this volume, and Helsvig, chapter 12, this volume.

128 Lefkaditou, chapter 1, this volume.

factors' that his tests identified were 'independent of knowledge and experience'.<sup>129</sup> Others imagined population-level bell curve distributions of pre-formed talent.<sup>130</sup> Such appeals to essential and knowable individual differences became particularly prominent during the trial period of curricular differentiation in advance of the nine-year reform.

There is, moreover, a notable similarity in the way both US and Norwegian reformers foregrounded hereditarian individualism at the height of their respective mid-twentieth century modernization efforts. In both cases, reformers advanced the notion that attention to natural individual differences could correct for inequalities associated with groups. In other words, just as hereditarian individualism was supplied as a race-neutral solution to the desegregation conflict in the US in the 1950s, so it was openly proposed as a class-neutral answer to socio-economic inequality in Norway at the same mid-century moment. As the older redoubt of early twentieth-century race science crumbled, the locus of discrimination (and where natural difference could be said to lie) was shifting from the level of the group (race or class) to the level of the individual. Claims of 'equality of educational opportunity' were reauthenticated in the process.<sup>131</sup> Thus, this particular reconstruction of intelligence appears to have served as an important regulating idea during periods of intensive modernization and re-equilibration in both the US and Norway. Of course, while measured individual intelligence would emerge as the colorless, classless, objective marker of worth, none of the methodological problems pertaining to hereditarian individualism explored here had actually been addressed in either national case.

Relatedly, race and class bias in test construction and interpretation remains an additional and compounding factor in these considerations. It is clear that in both the US and Norwegian cases, there were prolonged periods of early practice where tests were simply translated into the primary national language and not normed or standardized for the populations to which they were applied. Furthermore, Kyllingstad, Lefkadiou, and Håkon Caspersen's chapters in this volume all indicate that as IQ tests were standardized for a Norwegian majority, there was inadequate attention given to issues of ethnic, cultural, and linguistic diversity within Norway. This, of course, parallels developments in the US. Yet, this point

---

129 Helsvig, chapter 12, this volume.

130 Thue, chapter 11, this volume.

131 Porter, 'The Entanglement of Racism and Individualism'.

of similarity also marks an important discrepancy. Notably, the process of debate among US social scientists, beginning in the late 1920s, which unambiguously identified racial and class bias in IQ testing, appears not to have taken place in Norwegian expert discussion of test construction at any point over the course of the twentieth century. This gap in Norwegian expert conversation is the result, in part, of an unproblematic, tacit majoritarian understanding of Norwegian national identity as culturally homogenous.<sup>132</sup> It is also consistent with what Kyllingstad observes as a relative absence of critical race discourse in Norway, a kind of silence which can further facilitate a general perception of IQ as a fair, objective and color-blind measure of ability or intelligence.<sup>133</sup> As Kyllingstad and Caspersen both indicate, even late twentieth- and early twenty-first-century Norwegian revisions of the WISC-IQ remain inadequately standardized for ethnic minorities or immigrant groups within Norway. Kyllingstad also points to recent sociological analysis which has demonstrated that children of immigrants or those with minority-language backgrounds are disproportionately likely in Norway to be assigned to special education on the basis of IQ testing, and both he and Caspersen recount evidence of such ethnic and linguistic bias in current placement practices.<sup>134</sup>

Stepping back, this comparative analysis then reveals something more fundamental about where these two societies have marked out boundaries of inclusion and exclusion in their respective educational orders. Such boundaries have had a strong historical tendency to become barriers, and to correspond with the walls that run both within and around institutions. Over the twentieth century, Norway used IQ testing to draw a clear line between general and special education. The US had drawn such a line too, but then quickly set out to erect a much more expansive set of divisions across its public schools—delimiting types of education and vocation while also reproducing race and class—all according to so-called intelligence. Such comparative studies in educational boundary-making, then, also show how, under what pretexts, and to what degrees these two countries have chosen unequally and selectively to invest in what is more truly the vast, unmeasurable, and in an important sense unknowable potential of all their learners.

132 As explicitly taken up and discussed by Kyllingstad, chapter 10, this volume.

133 Jon Røyne Kyllingstad, 'The Absence of Race in Norway?', *Journal of Anthropological Sciences*, 95 (30 December 2017), 319–27, <https://doi.org/10.4436/JASS.95012>

134 Kyllingstad and Caspersen, chapter 14, this volume.

## Bibliography

- Birkemo, Asbjørn, *Pedagogisk-psykologisk rådgivning* (Oslo: Universitetet i Oslo, 2009).
- Bowles, Samuel, and Herbert Gintis, *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*. Reprint edition (Chicago: Haymarket Books, 2011).
- Brigham, Carl C., and Robert M. Yerkes, *A Study of American Intelligence, by Carl C. Brigham, ... A Foreword by Robert M. Yerkes* (Princeton: Princeton University Press, 1923).
- Carson, John, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750–1940* (Princeton: Princeton University Press, 2007).
- , 'Mental Testing in the Early Twentieth Century: Internationalizing the Mental Testing Story', *History of Psychology*, 17.3 (2014), 249–55, <https://doi.org/10.1037/a0037475>
- Chapman, Paul Davis, *Schools as Sorters: Lewis M. Terman, Applied Psychology, and the Intelligence Testing Movement, 1890–1930* (New York: New York University Press, 1990).
- Chauncey, Henry, 'Hearings before the Committee on Education and Labor, US House of Representatives 85th Congress, 2nd Sess. on Scholarship and Loan Program', § Committee on Education and Labor (26 February 1958).
- Clowse, Barbara Barksdale, *Brainpower for the Cold War: The Sputnik Crisis and National Defense Education Act of 1958* (Westport: Greenwood Press, 1981).
- Conant, James Bryant, Papers of James Bryant Conant, 1862–1987, 'Correspondence with William Alexander', 22 September 1958. UAI 15.898, Ability Grouping, Departmentalization, Box 108. Harvard University Archives.
- , *The American High School Today: A First Report to Interested Citizens* (New York: McGraw-Hill Book Company, 1959).
- Cravens, Hamilton, *Before Head Start: The Iowa Station & America's Children* (Chapel Hill: University of North Carolina Press, 1993).
- Day, John Kyle, *The Southern Manifesto: Massive Resistance and the Fight to Preserve Segregation* (Jackson: Univ. Press of Mississippi, 2014).
- Degler, Carl N., *In Search of Human Nature: The Decline and Revival of Darwinism in American Social Thought* (New York: Oxford University Press, USA, 1992).
- Douglas, Davison, *Jim Crow Moves North: The Battle over Northern School Segregation, 1865–1954* (New York, NY: Cambridge University Press, 2005).

- Desmond, Matthew, 'Capitalism,' *The 1619 Project: A New Origin Story*, ed. by Nikole Hannah-Jones and The New York Times Magazine (New York: One World, Random House Publishing Group, 2021).
- Fancher, Raymond E., *The Intelligence Men: Makers of the IQ Controversy* (New York: W.W. Norton & Company, 1987).
- Fendler, Lynn, and Irfan Muzaffar, 'The History of the Bell Curve: Sorting and the Idea of Normal', *Educational Theory*, 58. 1 (2008), 63–82, <https://doi.org/10.1111/j.1741-5446.2007.0276.x>
- Gardner, John, 'Observations on the Conant Study (Keppel's and Others' Feedback, Compiled and Edited by John Gardner).', n.d. Papers of James Bryant Conant, 1862–1987, UAI 15.898, High School March—June 1957, Box 42. Harvard University Archives.
- Goddard, H. H. 'Mental Tests and the Immigrant', *Journal of Delinquency*, 2 (1917), 243–77.
- Gould, Stephen Jay, *The Mismeasure of Man* (New York: W. W. Norton & Company, 2006).
- Howe, Michael J. A., *IQ in Question: The Truth about Intelligence* (London: SAGE Publications Limited, 1997).
- Hovde, Frederick, Hearings before the Committee on Labor and Public Welfare, U.S. Senate, 85th Congress, 2nd sess. on Science and Education for National Defense (28 January 1958).
- Kaestle, Carl, and Marshall Smith, 'The Federal Role in Elementary and Secondary Education, 1940–1980', *Harvard Educational Review*, 52.4 (1982), 384–408.
- Keller, Evelyn Fox, *The Mirage of a Space between Nature and Nurture* (Durham: Duke University Press, 2010).
- Kevles, Daniel J., *In the Name of Eugenics: Genetics and the Uses of Human Heredity* (Cambridge: Harvard University Press, 1995).
- Kliebard, Herbert M., *The Struggle for the American Curriculum, 1893–1958* (New York: RoutledgeFalmer, 2004).
- Kyllingstad, Jon Røyne, *Measuring the Master Race: Physical Anthropology in Norway, 1890–1945* (Cambridge: Open Book Publishers, 2014), <https://doi.org/10.11647/OBP.0051>
- , 'The Absence of Race in Norway?' *Journal of Anthropological Sciences*, 95 (30 December 2017), 319–27, <https://doi.org/10.4436/JASS.95012>
- Larsen, Lina Bryne, "'Med måling følger viden" Intelligensmåling i den norske folkeskolen 1900–1945' (Master's thesis, University of Bergen, 2007).
- Lazerson, Marvin, *Origins of the Urban School: Public Education in Massachusetts, 1870–1915* (Cambridge: Harvard University Press, 2013).

- Lippmann, Walter, 'The Mental Age of Americans', *New Republic*, 32.412 (25 October 1922), 213–15.
- , 'The Reliability of Intelligence Tests', *New Republic*, 32.414 (8 November 1922), 275–77.
- , 'Tests of Hereditary Intelligence', *New Republic* 32.416 (22 November 1922), 328–30.
- , 'A Future for the Tests', *New Republic*, 33.417 (29 November 1922), 9–10.
- Ludvigsen, Kari, and Åsmund Arup Seip, 'The Establishing of Norwegian Child Psychiatry: Ideas, Pioneers and Institutions', *History of Psychiatry*, 20.1 (1 March 2009), 5–26, <https://doi.org/10.1177/0957154X08089847>
- McDougall, William, *Is America Safe for Democracy?: Six Lectures Given at the Lowell Institute of Boston, Under the Title 'Anthropology and History, Or The Influence of Anthropologic Constitution on the Destinies of Nations* (New York: Charles Scribner's Sons, 1921).
- McNamara, Keith M., 'Testing Bias in Psychology, Law, and Public Policy, 1920–1980' (PhD Dissertation, University of Wisconsin-Madison, 2022).
- Minde, Henry, 'Assimilation of the Sami—Implementation and Consequences', *Journal of Indigenous Peoples Rights*, 3 (1 January 2005).
- Oakes, Jeannie, *Keeping Track: How Schools Structure Inequality* (New Haven: Yale University Press, 2005).
- Painter, Nell Irvin, *The History of White People*, Reprint (New York: W. W. Norton & Company, 2011).
- Paul, Diane B., *Controlling Human Heredity* (New Jersey: Humanities Press, 1999).
- Petrina, Stephen, 'The "Never-to-Be-Forgotten Investigation": Luella W. Cole, Sidney L. Pressey, and Mental Surveying in Indiana, 1917–1921', *History of Psychology*, 4.3 (2001), 245–71, <https://doi.org/10.1037/1093-4510.4.3.245>
- Porter, James Wynter, 'Constructing the "Gifted" and "Academically Talented" Student: "Intelligence," Intelligence Testing, and Educational Opportunity in the Era of Brown v. Board and the National Defense Education Act' (PhD Dissertation, Michigan State University, 2017).
- Porter, Jim Wynter, 'A "Precious Minority": Constructing the "Gifted" and "Academically Talented" Student in the Era of Brown v. Board of Education and the National Defense Education Act', *Isis*, 108.3 (1 September 2017), 581–605, <https://doi.org/10.1086/694446>
- , 'The Entanglement of Racism and Individualism: The U.S. National Defense Education Act of 1958 and the Individualization of "Intelligence" and Educational Policy', *Multiethnica*, 38 (December 2018), 3–17.

- , 'Guidance Counseling in the Mid-Twentieth Century United States: Measurement, Grouping, and the Making of the Intelligent Self', *History of Science*, 58.2 (1 July 2020), 191–215, <https://doi.org/10.1177/0073275319874977>
- Richards, Graham, *Race, Racism and Psychology: Towards a Reflexive History* (New York: Routledge, 2012).
- Rudolph, John L., *Scientists in the Classroom: The Cold War Reconstruction of American Science Education* (New York: Palgrave Macmillan, 2002).
- Rust, Val Dean, *The Democratic Tradition and the Evolution of Schooling in Norway, Contributions to the Study of Education* (Westport: Greenwood Press, 1989).
- Sandven, Johannes, *Intelligensmåling Ved Gruppeprøver* (Oslo, 1943).
- Sokal, Michael M., 'The Origins of the Psychological Corporation', *Journal of the History of the Behavioral Sciences*, 17.1 (1981), 54–67, [https://doi.org/10.1002/1520-6696\(198101\)17:1<54::AID-JHBS2300170108>3.0.CO;2-R](https://doi.org/10.1002/1520-6696(198101)17:1<54::AID-JHBS2300170108>3.0.CO;2-R)
- Spring, Joel, *The American School: From the Puritans to the Trump Era* (New York: Routledge, 2018).
- Stoddard, Lothrop, *The Revolt against Civilization: The Menace of the under Man* (New York: Charles Scribner's Sons, 1922).
- Telhaug, Alfred O., Odd Asbjørn Mediås, and Petter Aasen, 'The Nordic Model in Education: Education as Part of the Political System in the Last 50 Years', *Scandinavian Journal of Educational Research*, 50.3 (1 July 2006), 245–83, <https://doi.org/10.1080/00313830600743274>
- Terman, Lewis Madison, *The Measurement of Intelligence: An Explanation of and a Complete Guide for the Use of the Stanford Revision and Extension of the Binet–Simon Intelligence Scale* (New York: Houghton Mifflin Company, 1916).
- , *The Stanford Revision and Extension of the Binet–Simon Scale for Measuring Intelligence* (Baltimore: Warwick & York, Inc., 1917).
- Tyson, Carolyn, *Integration Interrupted: Tracking, Black Students, and Acting White after Brown* (New York: Oxford University Press, 2011).
- Urban, Wayne J., *More Than Science and Sputnik: The National Defense Education Act of 1958* (Tuscaloosa: University of Alabama Press, 2010).
- Wallin, J. E. Wallace, 'Who Is Feeble-Minded?' *Journal of the American Institute of Criminal Law and Criminology*, 6.5 (1916), 706–16, <https://doi.org/10.2307/1133344>
- Ydesen, Christian, Kari Ludvigsen, and Christian Lundahl, 'Creating an Educational Testing Profession in Norway, Sweden and Denmark, 1910–1960', *European Educational Research Journal*, 12.1 (1 March 2013), 120–38, <https://doi.org/10.2304/eeerj.2013.12.1.120>
- Zenderland, Leila, *Measuring Minds: Henry Herbert Goddard and the Origins of American Intelligence Testing* (Cambridge: Cambridge University Press, 2001).





## 14. IQ Testing, Education, Reification, and “Race” in Norway’s Social Democratic Welfare State

*Jon Røyne Kyllingstad and  
Håkon Aamot Caspersen*

---

In this chapter we take up and discuss a few examples of emerging and more overarching perspectives on the Norwegian history of IQ testing that appear throughout the book. These include reification and the pragmatic aspect of testing, as well as the relative absence of explicit discussions on “race”. We begin with comparative observations focusing on the Norwegian educational system and its connection to the social democratic welfare state, introduced by Jim Porter in the previous chapter.

### IQ Testing, the Unitary School and the Social Democratic Welfare State

The rise of IQ testing in Norway, as in many other countries, has been closely connected to the development of a mandatory public school. The Norwegian school system was characterized by the ideal of *Enhetskolen* [Unitary school]. This was an important precondition for the rise of IQ testing and the niche it came to inhabit in the Norwegian school system. Since the inception of universal schooling in 1889, the development of the Norwegian public school has been increasingly dominated by attempts to phase out a school system that channelled children of different backgrounds into educational tracks, thus reproducing social

inequalities. Instead, there was a push towards using the school as a tool to enhance national integration and social equalization, through a unified track that could give all students equal access to higher education. The unitary school was extended and reinforced through reforms that increased the length of the joint obligatory school from five years to seven years (1920), then to nine years (1969) and ten years (1997), along with an increasing equalization of the content and quality of schools throughout the country.

The increasing expansion of this unified system was paralleled by the growth of municipal and state institutions for children unable to cope with standard progression and teaching in the ordinary school. IQ testing was introduced and upheld as part of the education system mainly as a tool to identify these children. These testing practices were at first local and confined to cities, before the rise of a nationwide systematic development of special schools—and later the Educational Psychological Services (EPS) from the 1950s onwards—which led to the expansion of IQ testing practices.<sup>1</sup>

IQ testing in the Norwegian educational system has thus, historically, mainly been a method for drawing a shifting boundary between school children deemed “normal” educatable subjects and those who were not.<sup>2</sup> It has generally not been used for the division of “normal” pupils into different (vocational and academic) tracks within the ordinary school system, or to identify and offer special training to “gifted” children. A notable exception to this rule was the extensive group testing that was undertaken as part of the 1950s experiments with a tracked mandatory secondary school in preparation for the nine-year school reform. These mass-testing experiments were terminated, however, when plans for a differentiated secondary school were abandoned and the nine-year unitary school was enacted.

These aspects of the Norwegian history of IQ testing in the school system contrast, as described by Porter, with the United States, where IQ testing penetrated earlier and more deeply into the school systems. In the US, such testing became a tool to not only draw the boundary between

1 Asbjørn Birkemo, *Pedagogisk-Psykologisk Rådgivning* (Oslo: Unipub, 2009).

2 Bjørn Hamre, Tine Frisrup and Gerd Christesen, ‘The subject of exemption: through discourses of normalization and individualization in Denmark’, *Nordic Journal of Social Research*, 7.2 offers an interesting comparative discussion of normalization and the social construction of the educable subject in neighboring Denmark.

normal and weakly gifted children, but also between ranked categories of children in the ordinary school. Porter explains this by pointing to the development of systems of organizationally differentiated education adapted to the assumed abilities of the individual child, often based on IQ testing.

The rapid expansion of IQ testing in US schools was also the product of a mutual interaction between the profession of psychology as it worked to develop its legitimacy and social relevance, test publishers and individual psychologists who made good money from the emergence of a test industry, and school administrators in search of tools to drive tracking practices in schools. Norway, in contrast, only saw the establishment of psychology as a profession after World War Two, and never built a profitable test industry at the US level, since the nation, and thus the test market, was too small and less socially diverse. In Norway, the commercial aspect became an obstacle rather than a driving force for test development, and continues so today.

Norwegian adaptations of WISC tests are a case in point. The Norwegian WISC-R (1978) was funded by the state, through the Norwegian Research Council. However, in the 1990s, The Norwegian Psychologists Association was not able to get public funding for a Norwegian WISC-III and found it difficult to finance an adaptation based only on test sales. They decided to sell their test unit to the private company Assessio, which launched a Norwegian version of WISC-III in 2003. But this version had not undergone a thorough adaptation, and its quality was soon questioned, as noted in the introduction to this volume. The Norwegian standardization of WISC-IV (2009), in turn, was undertaken by Pearson Assessment, the present owner of the Wechsler tests. By their own admittance, the resources put into the Norwegian version was dictated by the (small) size of the Norwegian market. Today Pearson considers the Nordic countries as one market, and the present Norwegian version, WISC-V, is normed on a Nordic population sample.<sup>3</sup>

The commercial aspect of the tests points towards a more fundamental

---

3 See Tor Herman Andreassen, 'WISC-III og de eldste ungdommene: praktiske erfaringer', *Tidsskrift for Norsk psykologforening* 43, no. 5 (2006), Nina Strand, 'WISC-III skaper uro', *Tidsskrift for Norsk psykologforening* 42, no. 6 (2005), 519–22, David Wechsler, *WISC-IV. Norsk versjon Manual del 1* (2009), pp 7–8, *WISC-V Norsk versjon Manual Del 1* (2017).

dissimilarity in the Norwegian context of IQ testing as compared to the US, namely a significant difference in their political and societal development—free market capitalism in the US, with its weak labor movement and relatively weak social welfare systems, versus Norway with a much stronger labor movement and the rise of a social democratic welfare state. According to Porter, industrialists and corporate interests, along with racism and classism, played a significant role in the rise of educational tracking in the US. The influence of big capital seems to have been far less important in Norway.

The rise of the Norwegian unitary school was, from the 1880s until around 1920, driven mainly by the dominating social liberal and national democratic *Venstre* movement, characterized by ideals of social, cultural, and national integration; and from the 1930s onwards, mainly by the social democratic Labor party. Similar to the US, school reforms and the development of testing, as well as a scientific approach to teaching and school development, were intertwined with the rise of pedagogy as a science and by the reform pedagogy movement. In contrast to the US, however, the Norwegian educational field was characterized by alliances between Labor party educational politicians and progressive educationalists within academia and the school system from the 1930s to the 1970s. In this volume, we have encountered a couple of examples of testing in Norwegians schools that took place outside the realm of special education—group testing of first-graders in Oslo from the 1930s to the 1960s, and experiments with tracked secondary schooling in the 1950s. In both examples, ability testing was seen as a tool to adapt teaching to the abilities of the individual pupil within the frame of the unitary school, and both examples were related to progressive educational ideas associated with the Labor party and aimed at social equality and national integration.

As pointed out by Thue in chapter 11, this may help explain some aspects of the IQ controversy that arose in the late 1950s, where much of the criticism came from a culturally conservative or liberal, non-socialist point of view, directed against the rule of experts and public authorities. With regard to actors and topics, the controversy on the IQ testing of Oslo schoolchildren overlapped partly with the media debate a couple of years previously, triggered by a court case about forced sterilization, IQ testing, and psychiatric assessments, as discussed by Haave and

Kyllingstad in chapter 4. Both disputes addressed questions of how the interference of public authorities in an individual's life should be regulated and delimited. Who had the competence to set this limit, and what role should intelligence tests have? It could be argued that both disputes highlighted inbuilt tensions typical of a social democratic welfare state that intervened strongly in social structures affecting the individual lives of members of Norwegian society.

These controversies took place when the paternalistic and technocratic attitudes that had characterized the treatment of people with an intellectual disability began to meet increasing critique, and both cases helped to bring into place new practices that were based on the acknowledgment of the integrity and rights of the individual. An increasing focus on individual rights did not, however, lead to a reduced use of IQ testing. As Haave and Kyllingstad argue regarding involuntary sterilizations, and as Skålevåg and Gröning also discuss regarding the Criminal Act's definition of legal unaccountability, an increased focus on rights could lead to an increased reliance on IQ testing as an “objective” method to adjudicate access to these rights.

### The Relative Absence of “Race” and Ethnicity in Norway

The Norwegian IQ testing controversies in the late 1950s were multifaceted and principled, and included test-related aspects well beyond the concrete issue that had sparked the debate. The critique did not, however, touch upon issues of ethnic, racial, or even class bias. This is interesting to note, since controversies about these issues have surrounded IQ tests since their infancy, and because much existing historiography on IQ testing addresses questions about racial and ethnic bias and discrimination. The relative absence of issues of “race” and “ethnicity” is discussed by both Kyllingstad and Porter in chapters 10 and 13, but it is still worthy of a few additional words that add to and complement what has been noted so far—the relative absence of discussions about race does not mean that some of the historically associated problematic usages and consequences of IQ testing did not exist.

Intelligence testing was initially developed mainly as a tool to identify individuals with a “mental deficiency”, and the attitude towards these individuals was strongly affected by eugenic thinking. Eugenic thinking in interwar Norway had, like its US counterpart, strong elements of racial thinking and valuations of “superiority” and “inferiority”. Influential Norwegian eugenicists advocated, for example, a sterilization policy directed against one specific ethnic group, the Romani. They failed to pass a law that specifically targeted this group. Yet even so, the law was enacted in a way that discriminated against the Romani, and IQ testing played a role in legitimizing this discrimination.<sup>4</sup>

In 1927, the medical student Mikael Kobro IQ tested seventy-one ‘vagrant children’ using a non-normed Norwegian translation of Stanford–Binet and found that almost eighty-six percent were ‘more or less underachievers’. This led him to conclude that the Romani were ‘an inferior people’ whom he, with reference to Lewis Terman’s research, placed ‘intellectually, in a class with Negroes, Indians and Mexicans’. His study was commissioned by the Christian philanthropic association *Norsk misjon blant hjemløse* [Norwegian mission among the homeless], responsible for much of the practical implementation of the state’s Norwegianization policy to combat the culture of the Romani through forced settlement and forced custody of Romani children, and Kobro’s study contributed significantly to their decision to advocate eugenic sterilization as a supplement to this policy. Relative to the size of the population, Romani people (mainly women) were more often sterilized than the rest of the population. This discriminatory practice was related to a far-reaching and ruthless policy of Norwegianization and a derogatory view of the Romani way of life and culture, and it is likely that Kobro’s study contributed to this discrimination.<sup>5</sup>

There may be other examples of minority people being intelligence tested because they belonged to a minority, and based on notions of racial differences in hereditary intelligence. The general pattern, however, still seems to be that IQ was linked to issues of race and ethnicity to a far lesser extent than in the United States. In the early decades of the twentieth

4 See Per Haave, *Sterilisering av tatere 1934–1977: en historisk undersøkelse av lov og praksis* (Oslo: Norges forskningsråd, 2000).

5 Per Haave, ‘Sterilisering av tatere - kirurgi på “rasemessig” grunnlag?’, in *Romanifolket og det norske samfunnet: Følgene av hundre års politikk for en nasjonal minoritet*, ed. by Bjørn Vinden (Bergen: Fagbokforlaget, 2000), pp. 39–57.

century, for example, relatively extensive physical-anthropological race research was carried out in Norway. Large population samples were racially classified through studies of external physical characteristics to explore theories about the racial origins of Sámi and non-Sámi Norwegians. Even local populations of majority Norwegians were understood to be racially different from each other. Although this race research was characterized by notions of mentally superior and inferior “races” and was partly related to racist eugenic thinking, it was never directly linked to IQ testing or related psychometric investigations in Norway, as was the case in the US and elsewhere.<sup>6</sup>

In connection to this, it is important to note that other minorities in Norway, such as the Sámi and the Kven, were subjected to an assimilation policy aimed at eradicating their culture. Thus, even if Rom, Romani, Sámi, and Kven people were commonly portrayed as racially different and intellectually inferior in public and academic discourse, this attitude was combined with a cross-party policy of Norwegianization—they should be transformed into “Norwegians”. Importantly, the public school, with its aim of national integration, was the key instrument for this project, which was only officially terminated in 1963, and even later for the Romani, in the 1980s.<sup>7</sup>

Regarding the concept of race, Norway followed the same pattern as most other countries in Western Europe in the decades after World War Two. It became less and less common to conceptualize cultural or linguistic diversity within the Norwegian population as racial differences. A growing opposition to biological race thinking did not lead to the adoption of a social constructionist notion of “race” in Norwegian society. Instead, race became a “taboo” word, substituted by ethnicity,

---

6 Jon Røyne Kyllingstad, *Measuring the Master Race: Physical Anthropology in Norway 1890–1945* (Cambridge: Open Book Publishers, 2014) and Jon Røyne Kyllingstad, *Sannhets og forsoningskommisjonen. Ekstern utredning. Undersøkelse av fysisk antropologisk forskning på samer, kvener og skogfinner med vekt på norske myndigheters rolle Del 2* (Oslo: Stortinget, 2022).

7 The historical assimilation policy towards the Romani was documented and assessed by a government-appointed committee in the report NOU 2015: 7 *Assimilering og motstand—Norsk politikk overfor taterne/romanifolket fra 1850 til i dag*. The assimilation policy towards Sámi, Kven and Forest Finns was documented and assessed by the Norwegian Truth and Reconciliation Commission’s report *Dokument 19 (2022–2023) Sannhet og forsoning - grunnlag for et oppgjør med forsoningspolitikk og urett mot samer, kvener/norskefinner og skogfinner*. In practice, in schools, the Norwegianization policy continued for a long time after it had been abolished in principle.



while the idea of Norway as an ethnically homogeneous country was reinforced.<sup>8</sup> It is against this background that we need to understand the conspicuous absence of the issue of linguistic and cultural minorities in the two Norwegian postwar test adaptation projects *Norsk standard av Terman-Merrills Stanford-revisjon av Binet-prøvene* (1954) and *WISC-R* (1978), and that none of the test manuals addressed how the test could or could not be used on children who did not belong to the group for which the test was adapted and normed.

This omission is particularly striking with regard to the *WISC-R*, since the original American *WISC-R* test had been created with the US “IQ war” about intelligence and race as a backdrop—and since the Norwegian version was created at a time when the Norwegianization policy had just been replaced by a policy of preserving Sámi language and culture, and when immigration was beginning to make its mark on the composition of the Norwegian population. It is therefore notable that it was only in the early 2000s that this issue was put squarely on the public and professional agenda, although not on the basis of “race” but of language and ethnic background.

It is tempting to propose that this can be understood through specific aspects of the Norwegian context: a long history of regarding Norway as a homogeneous nation, along with an ideology of color-blindness and a tradition of cultural assimilation. Though officially abandoned, this way of thinking still had repercussions well into the 1990s, when notions of imagined sameness also constituted invisible fences.<sup>9</sup> On the other hand, the overrepresentation of minority children in special education, as well as misinformed IQ testing of minority children, are far from being solely Norwegian phenomena, and have proved widespread in many Western countries. In fact, the studies by Pihl and Aagaard, which exposed such problems in the EPS, were inspired by international research that demonstrated and confirmed that the situation in Norway was not very different from other countries.<sup>10</sup>

8 Jon Røyne Kyllingstad, ‘The Absence of Race in Norway?’, *Journal of Anthropological Sciences* JASS Forum 95 (2017), p. 319.

9 Marianne Gullestad, ‘Invisible Fences: Egalitarianism, Nationalism and Racism’, *Journal of the Royal Anthropological Institute*, 8 (2002).

10 Knut Erik Aagaard, *Den språklige faktor: pedagogisk-psykologisk utredning av barn med minoritetsspråklig bakgrunn*. dr.dissertation (Oslo: University of Oslo, 2010), pp. 123–159 and Joron Pihl, *Etnisk Mangfold i Skolen: Det Sakkyndige Blikket*. 2. utg. (Oslo: Universitetsforlaget, 2010).

## The Practical Reification of IQ

IQ tests have often functioned or stood in as more neutral measuring instruments in different assessment procedures involving “professional judgment”. Such use of IQ testing suggests that the tests have, to a greater or lesser extent, become “black boxed”, meaning that the normative, ontological, and epistemological premises on which the tests are based are made invisible. The tests thus come to appear as neutral instruments that produce objective knowledge, provided that they are handled by experts who ensure that the technology works as it should, much like Hacking describes the ‘looping’ effects of other scientific classificatory practices.<sup>11</sup> This has contributed to the reification of the IQ tests’ concept of intelligence: Intelligence has, in this sense, been transformed into a thing, a social fact that exists in itself, and that can be measured. In line with such thinking, one can say that a person who acts intelligently (for example by scoring high on an IQ test) has a lot of “intelligence”. The reified term is then, in the next round, often turned into an explanation through circular reasoning—in the case of intelligence, the idea of intelligence as an innate quality can explain the more or less intelligent behavior.

The history of early IQ testing is often critically narrated as a story about the conscious reification of intelligence, undertaken by individuals with specific (bad) intentions, such as confirming a racist and classist worldview. This, in turn, bolsters racist and class-based social structures or, importantly, enhances the prestige, legitimacy and expert power of the professions that claim to be able to measure and control the thing called intelligence. Steven Jay Gould’s *The Mismeasure of Man* is a very influential and much-cited example of such an interpretation of the history of IQ testing.

Many of the preceding chapters in this book tell a somewhat different story. They can indeed be read as narratives about how IQ testing has led to reification, but this has often not been the result of explicit intentions, struggle for expert power, scientific hubris, or ideological conviction. Instead, it has, in many cases, been the result of the practical functions

---

11 Ian Hacking, ‘The looping effects of human kinds’, in *Causal Cognition: A Multidisciplinary Debate*, ed. by Dan Sperber, David Premack, and Ann James Premack (Oxford: Oxford University Press, 1996).

and use of tests. Within the test field, there have often been reservations about the explanatory power of test results, and disagreement and ambiguity about what the test measures. Nevertheless, the tests have also, in many cases, helped to produce the concept of intelligence as a thing.

Gröning and Skålevåg, for instance, convincingly argue in chapter 3 that intelligence tests have contributed to a reification of intelligence in criminal law. They explain this not as a product of professional strategies to enhance the legitimacy and authority of the psychiatrist, but rather as a result of the role that intelligence tests have played in the interaction between forensic psychiatry and law in assessing the (un)accountability of intellectually disabled offenders. Intelligence testing, in short, helped transform the mind of the unaccountable offender into a measurable thing. Despite both psychiatrists and lawyers regularly expressing skepticism and reservations about the value of IQ scores, the tests came to legitimize the objectivity and reliability of forensic psychiatric expert opinions. From these assessments, they also found their way into the Penal Code as a tool to clarify the legal delimitation of unaccountability, and thus help give lawyers and judges the authority to deal with difficult issues. More than just being driven by power struggles, the reification of intelligence was thus ultimately the result of an interaction between law and psychiatry, and a product of practical needs to solve a difficult problem in the context of law: drawing a line between unaccountable and accountable offenders.

This process has similarities to other processes encountered in this book. The introduction of IQ testing did, undoubtedly, contribute to allocating legitimacy and expert power between disciplines and professions, while the incorporation of such technology into existing expert knowledge regimes and institutionalized processes simultaneously contributed to the test's legitimacy. It is, however, difficult to explain the entry of tests into Norway solely as a story of professional struggle to gain legitimacy and authority. In Norway, it was initially psychiatrists more than psychologists who initiated the use of IQ testing (in close collaboration with pedagogues). Even though it may have been the case, as Ydesen et al note in chapter 2, that the idea of intelligence as a hereditary and stable thing had stronger support among Scandinavian psychiatrists than among educators, the early

history of testing in Norway was characterized by a close cooperation between psychiatrists, reform-oriented pedagogues, and psychologists, and not necessarily by disciplinary conflict or a "battle of professions".

The story that we have seen here is not that of psychiatry, armed with an objective measurement tool, entering a domain previously controlled by teachers and taking control over assessments that had previously been based on the teachers' professional judgement. On the one hand, the value of test scores was partly justified by the argument that they coincided with the teachers' discretionary assessments of the student. On the other hand, the tests never worked as objective sorting machines. The cut-off points between different categories of pupils were only indicative; in practice, the tests were part of a broader assessment where both the psychiatrist's and the teachers' discretionary judgment was taken into account, as well as conditioned by practical, organizational, financial, and social factors.

Furthermore, the relationship between IQ testing and expert power is not clear-cut and simple. On the one hand, to the extent that a group, such as psychiatrists or psychologists, can claim to be the only ones with the knowledge and skills necessary to administer a test and interpret the test results, this group gains control over assessments that include testing. But this is also a double-edged sword. If the test score is perceived as a neutral and objective measure, it also limits the scope for the expert's exercise of discretionary judgment. The changed policy of the Norwegian Sterilization Board in the wake of the court case of the involuntarily sterilized woman in the 1950s testifies to this. A more restrictive practice of granting applications for involuntary sterilization of "mentally retarded" individuals, together with a desire to avoid new court cases, meant, by all accounts, that the importance of the psychiatrist's discretionary assessments was downplayed while the apparently objective IQ score was given greater weight. This means, in a sense, that more power was given to the test and its in-built ontological, epistemological and normative premises, while less power was given to the individual psychiatrist's observational and qualitatively-oriented expert judgment.

When it comes to tests that can be administered and interpreted mechanically by people without special training, as was the case with group tests of intelligence, the relationship between expert power and

tests becomes even more complex. On the one hand, the tests served as a counterweight to expert power, since anyone could buy the tests and ordinary teachers could make use of them. On the other hand, this is perhaps the foremost example of the black boxing of IQ test technology and reification of intelligence. To the extent that teachers used the tests and test results mechanically, without necessarily having insight into the ideas, premises, and assumptions the test was based on, we can here say that the test developers' expert power was exercised indirectly and at a distance every time the tests were used. Although perhaps more clearly illustrated by group testing, this is also a prevailing issue when it comes to individual IQ tests such as WISC. As indicated in the following reflections from Petra, an MA student in special needs pedagogy whom we met in chapter 7, upon recently completing the WISC-V certification part of the course:

I am of course not a statistician nor a neuropsychologist. And I am thinking I do not need to understand everything, but I need to know how to operate the tool. To know that it is built on science, and that A leads to B.

Petra added that one of the most important things learned through the course was the practice of following and consulting the manual.

The relationship between disciplines, expert power, and intelligence testing seems to have changed from the 1950s onwards. The establishment of university education in pedagogy and psychology, and the emergence of psychology as a well-organized profession may have laid the groundwork for a more explicit struggle about control in the field of intelligence testing. The public IQ test controversies of the late 1950s, however, meant that academic discussions about the reliability and validity of IQ tests and about what is meant by the term intelligence, was moved out of a limited professional sphere and turned into a public debate that took place in mainstream newspapers and even in Parliament. The lay criticism of IQ testing was generally directed against an unspecified group of psych-experts, often referred to as "psychologists". As Emberland notes, in Norway, strong associative links between the psychologist and testing were cemented in the popular imagination. Leading psychologists, in turn, defended IQ testing, by arguing that the problem was not the tests in themselves, but bad and unqualified testing by psychiatrists and ordinary teachers who lacked

the necessary psychological insight and psychometric competence for administering and interpreting such tests. They maintained that only psychologists and university-educated pedagogues were competent to conduct IQ testing.

The general message from these leading psychologists seems to have been that tests had an important function, but only if they were used correctly, and that only university-trained pedagogues and psychologists could do this. At the same time, it is interesting to note that their arguments in the public discussion addressed core scientific issues, such as the relationship between nature and nurture and the stability of intelligence (as measured by IQ) through an individual's life course. If we apply the black box metaphor to this debate, it appears that the psychologists' arguments to defend their own test-related expert power included the opening of the IQ black box and the public discussion of its contents, and that this may have helped to increase their expert power in this field, rather than weaken it.

These public controversies took place at a time when a nationwide expansion of school psychology offices was underway, creating an important job market for graduates from the new university studies in education and psychology. The question of the division of labor between medicine, pedagogy, and psychology in these offices now became a key issue. In this context, it is interesting to note the strong test-psychological orientation of the emerging university discipline of pedagogy, and that the psychologists involved in the IQ-related controversies actually described pedagogy as a psychological discipline on a par with psychology when it came to IQ testing. However, in the longer run, it became clear that in Norway, too, it was the psychologists proper who came to gain jurisdiction over the test field. This becomes apparent if we look at who is behind the Norwegian versions of key IQ tests.

The Norwegian standardization (1931) of the original Stanford–Binet (1916) was carried out by a psychiatrist with close ties to the teaching profession and reform pedagogy. The next version (from 1937) was transformed into a Norwegian version (1954) by pedagogy students, under the leadership of a professor of pedagogy at the Pedagogical Research Department at the University of Oslo. It was published and marketed by this department in collaboration with the book publisher Cappelen. The Norwegian versions of the Wechsler tests (WAIS,

WPPSI 1974, and WISC 1978) were developed by psychologists and published and marketed by the Norwegian Psychological Association (in agreement with The Psychological Corporation), and it was the Norwegian Psychological Association who decided that these tests should only be sold to psychologists and specific university-educated pedagogues who had been test-certified. When looking to arenas where IQ testing is taught today, one sees that it is predominantly psychologists who teach, and pedagogues who are taught in education programmes that offer test certification. However, looking beyond the confines of clinical psychology and psychiatric services, it is the pedagogues who are likely to carry out the bulk of the testing in the Educational Pedagogical Services. This is a service that for many educated Norwegian psychologists today seems somewhat peripheral—of a lower professional status than working in the health sector that offers clearer paths to further professional development.

### Final Remarks

The history of IQ testing is full of paradoxes. IQ testing has been controversial throughout most of its history. From the outset, both scientific and ideological criticism has been directed at the implicit and explicit normative, ontological, and epistemological premises on which IQ tests are based, and the problematic consequences of such tests on society and for individuals. These controversies have struck at the very core of what intelligence tests are. Within the professional communities that have developed, used, and defended the use of IQ tests, there have been disagreements and shifting views about what it is that IQ tests actually measure. This ambiguity has helped make IQ a flexible concept, which is probably a contributing factor in its success. The malleability of IQ and IQ tests has made it possible for both to move across national borders and into different social contexts, where they have been adapted to different societal needs, ideals, practical conditions, and scientific theories, and fulfilled a number of different societal functions. Along the way, IQ tests and related ideas and practices have been shaped by—and have helped shape—the social and cultural conditions of which they have become part.

On the other hand, while IQ tests are flexible and adaptable, much of their legitimacy rests on the notion that they, in a certain sense, are stable and unchangeable. An IQ test produces a clear measure in the form of an IQ score. The legitimacy of an IQ score rests on the idea that an IQ test always measures the same things as another IQ test. The legitimacy of IQ tests is further protected by professional expertise. In a paradoxical way, the history of IQ testing has been characterized by both the ambiguity and adaptability of IQ, and the notion of IQ as a stable and unchangeable measure of an objectively existing thing called intelligence.<sup>12</sup> It may be argued that the mutually reinforcing but changing interplays between black boxing and controversy, adaptability, and stability have contributed significantly to the spread of IQ testing across national borders, shifting societal contexts, and scientific justifications, and helped maintain its existence across such a long time span.

## Bibliography

- Aagaard, Knut Erik, *Den språklige faktor: pedagogisk-psykologisk utredning av barn med minoritetsspråklig bakgrunn* (PhD dissertation, Oslo: University of Oslo, 2010).
- Andreassen, Tor Herman, 'WISC-III og de eldste ungdommene: praktiske erfaringer', *Tidsskrift for Norsk psykologforening*, 43.5 (2006).
- Asbjørn Birkemo, *Pedagogisk-Psykologisk Rådgivning* (Oslo: Unipub, 2009).
- Byford, Andy, 'The Mental Test as a Boundary Object in Early-20th-Century Russian Child Science', *History of the Human Sciences*, 27.4 (2014), 22–58, <https://doi.org/10.1177/0952695114527598>
- Gullestad, Marianne, 'Invisible Fences: Egalitarianism, Nationalism and Racism', *Journal of the Royal Anthropological Institute*, 8 (2002), 45–63, <https://doi-org.ezproxy.uio.no/10.1111/1467-9655.00098>

---

12 Arguments that parallel and find support in Andy Byford's analysis of the history of mental testing in Russian child science, that he argues was characterized by a 'peculiar dynamic' of '(de)legitimation' as it 'swung between black boxing and instrumentalization on the one hand, and scandal and controversy, on the other'. See Andy Byford, 'The Mental Test as a Boundary Object in Early-20th-Century Russian Child Science', *History of the Human Sciences*, 27.4 (2014), 22–58 (p. 26).



- Haave, Per, 'Sterilisering av tatere - kirurgi på "rasemessig" grunnlag?', in *Romanifolket og det norske samfunnet: Følgene av hundre års politikk for en nasjonal minoritet*, ed. by Bjørn Vinden (Bergen: Fagbokforlaget, 2000).
- , *Sterilisering av tatere 1934–1977: en historisk undersøkelse av lov og praksis* (Oslo: Norges forskningsråd, 2000).
- Hamre, Bjørn, Trine Fristrup and Gerd Christesen, 'The subject of exemption: through discourses of normalization and individualization in Denmark', *Nordic Journal of Social Research*, 7.2 (2016).
- Kyllingstad, Jon Røyne, *Measuring the Master Race: Physical Anthropology in Norway 1890–1945* (Cambridge: Open Book Publishers, 2014), <https://doi.org/10.11647/OBP.0051>
- , *Sannhets og forsoningskommisjonen. Ekstern utredning. Undersøkelse av fysisk antropologisk forskning på samer, kvener og skogfinner med vekt på norske myndigheters rolle Del 2* (Oslo: Stortinget, 2022).
- , 'The Absence of Race in Norway?', *Journal of Anthropological Sciences*, JASS Forum 95 (2017).
- Pihl, Joron, *Etnisk Mangfold i Skolen: Det Sakkyndige Blikket*. 2. utg. (Oslo: Universitetsforlaget, 2010).
- Strand, Nina, 'WISC-III skaper uro', *Tidsskrift for Norsk psykologforening*, 42.6 (2005).
- Wechsler, David, *WISC-IV. Norsk versjon Manual del 1* (Pearson, 2009).
- , *WISC-V. Norsk versjon Manual Del 1* (Pearson, 2017).

# Afterword

*Annette Mülberger*

---

Readers arriving at this point will probably feel overwhelmed by everything they have read. *Historicizing IQ Testing* has much to offer; there is surely something for everyone amongst these fourteen histories about the arrival, the circulation, and the various uses of intelligence tests in Norway.<sup>1</sup> The first chapters take the reader directly into the educational environment. The arrangement of the chapters is roughly chronological, each exploring a different period and area in which the test played a role (education, law, clinics, specialized institutions, etc.). There is, consciously, no unity. The manifold nature of the collection is not helpful when seeking common ground. At the same time, it makes the reading extremely exciting—every chapter a new adventure!

It is a privilege to be guided through easy understandable chapters that delve into Norway's history and cultural idiosyncrasies, a context that—due to the language barrier—often feels far away.<sup>2</sup> Yet, the chapters reveal how the north is much closer to Europe and the United States than might be expected. As in other places, the test(s) arrived at a time when unitary schooling was being discussed. Within the institutional setting of the 1920s and after, an “objective” selection of children for special education programmes and mental disability institutions seemed fitting.

---

1 In a few chapters, the developments in other Scandinavian countries such as Denmark and Sweden are considered.

2 At the ESHHS we had some contact with these histories, thanks to the work of our esteemed and now retired colleague, the psychologist Karl Harvor Teigen. He published a textbook on the history of psychology in Norway: Karl Harvor Teigen, *En Psykologihistorie* (Oslo: Fagbokforlaget, 2015 [2004]). In relation to the history of IQ testing, see also his article on Thomas Parr in Karl Harvor Teigen, 'A note on a Norwegian precursor of Charles Spearman', *Journal of the History of the Behavioral Sciences*, 22.1 (1986).

All this should not lead us to think that the introduction of testing was easy. Also, in Norway, IQ tests were received by some with skepticism and critique (chapter 3 by Skålevag and Gröning about the legal history is especially telling). Similar to other contexts, the adaptation and standardization of tests to the local population and culture turned out to be rather “thorny” (to use Lefkaditiou’s expression in chapter 1).

At the same time, the reader’s sense of familiarity is soon mixed with excitement at each encounter with some outstanding aspects in these northern accounts. The skeptical attitude shown by the psychiatrist Dahlstrøm (and his emphasis of the environment when explaining criminal behavior), Reymert’s aborted mission in adapting the Stanford–Binet, the heated debates around sterilization, or the testing of school children in the 1950s in Oslo, are just a few of the highly valuable insights provided by the book.

Above everything else, the reader gets the impression that, despite some tensions, IQ testing appeared as a medium for cooperation among professionals. Thus, chapter 2 by Ydesen, Hovland and Vikstrøm observes a ‘productive cooperation’ between psychologists and special schoolteachers. In a similar vein, in chapter 14, Kyllingstad and Caspersen conclude: ‘The early history of testing in Norway was characterized by a close cooperation between psychiatrists, reform-oriented pedagogues, and psychologists, and not necessarily by disciplinary conflict or “battle of professions”’. This is remarkable! In my research on the introduction of intelligence testing in Spain, I came across hot-tempered attacks and competitive fights between psychiatrists and psychologists on the one hand, and pedagogues, p(a)edologists, and teachers on the other. At this point, I can only speculate that, beyond the Mediterranean temperament, the difference certainly must have an explanation (probably related to the institutional setting).

Overall, the book offers new material and valuable insights. Its content is truly inspiring. In the following, I want to share two final reflections that came to my mind while reading. Firstly, about the center-periphery dichotomy. The editors use it without any definition. Early northern testers sought training in the United States, importing the Stanford–Binet into their home context. It seems natural for a historian to try to find a way to conceptualize how the local testers situated themselves with regards to the places they perceived as “centers” within

this imbalanced power distribution (I, at least, used the terminology in that way).<sup>3</sup>

Yet, it is one thing to acknowledge an actor's perspective, and another to situate one's own historical account on the periphery. This seems to be the case in Porter's chapter 13, offering a comparison between the developments in the United States and Norway. Thereby, the nature-nurture debate and other themes dominating the American narrative acquire a role as 'baseline', through which to assess and judge the history of IQ testing in the north. In this light, the undefined category of 'Norway' appears as a 'slowly' developing society, in which the test gained only 'limited' social impact (It might be of interest here to add that, if we look at other European countries, Norway was not an exception).

Such comparisons are not rare. Twenty years ago, a historical case study by Mulder and Heyting compared the way intelligence testing was done in the Netherlands with the developments in the United States. They concluded that the Dutch use of tests was rather 'pragmatic'; not 'generating' the heated debates 'found in other cultures'.<sup>4</sup> The move seems clever: the Anglo-Saxon account of 'the IQ controversy' is nearly omnipresent nowadays, not only in psychology textbooks, but also on social media. Linking a more "remote" history to a well-known account seems like a good strategy for a text to reach a wider readership.

Yet, such a move tends to further (artificially) enlarge the role of the "center", eclipsing histories from other "peripheries". It should be added, though, that the authors of *Historicizing IQ Testing* made a highly valuable effort to go beyond national boundaries by examining intelligence testing in several Scandinavian countries. This is already more than what we are used to. Still, one might wonder whether the history of IQ testing presented in these chapters might have some similarities or connection to initiatives and dynamics in the Netherlands, Finland, Austria, Hungary, or France. Therefore, we hope that the volume will trigger more trans- and cross-national studies.

---

3 Annette Mülberger, 'Appropriation of psychological testing in the Spanish pedagogical context', in *The Circulation of Science and Technology: Proceedings of the 4th International Conference of the ESHS*, ed. by Antoni Roca-Rosell (Barcelona: SCHCT-IEC, 2010), p. 626.

4 Ernst Mulder and Frieda Heyting, 'The Dutch Curve: The introduction and reception of intelligence testing in the Netherlands, 1908–1940', *Journal of the History of the Behavioral Sciences*, 24.4 (1998), 349.

Furthermore, any national comparison raises the question about national(ist) competition and the author's situatedness. What does it mean if testing started "earlier"? After all that we know today, would we still take this as something positive and as a sign of "advancement"? Does it make sense to take the development of one country as "the norm" by which to judge the histories of other societies or places? Wouldn't this mean reproducing precisely one of the main points of critique of the "normalizing" IQ test?

Secondly, the current volume is timely. It is the outcome of at least four interesting historiographic turns that took place during recent decades:

1) Within the historiography of psychology, a formerly exclusive interest in laboratory science (cultivated by Boring<sup>5</sup> and his followers) has changed since the 1980s, making room for a study of what was sometimes called "applied psychology". It reflected a more general trend in the history of science, blurring the boundaries between science and knowledge that upgraded the standing of professional expertise.

2) Within the history of science, a second shift from theory to scientific instruments and practice took place, stirring an interest (since the 1990s) in the role of psychological instruments. Inspired by works such as Rheinberger's<sup>6</sup>, historians realized that instruments are not neutral, passive tools but theory-loaded, playing a significant social and professional role (see, e.g., Gundlach's diagnosis of the 'totemic' role of the Hipp chronoscope and Gigerenzer and Sturm arguing that psychologists' familiarity with statistics and computers inspired their theoretization<sup>7</sup>).

*Historicizing IQ Testing* follows this trend, dealing not only with testing but also with the test as tool and material object. For example,

---

5 Edwin (Garrigues) Boring, *A History of Experimental Psychology* (New York: Appleton Century-Crofts, 1950 [1929]).

6 Hans Jörg Rheinberger, *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube* (Stanford, CA: Stanford University Press, 1997).

7 Horst Gundlach, 'The Hipp chronoscope as totem pole and the formation of a new tribe', *Teorie & Modelli*, 1.1 (1996), 65–85 and Gerd Gigerenzer and Thomas Sturm, 'Tools = Theories = Data? On some circular dynamics in cognitive science', in *Psychology's Territories: Historical and Contemporary Perspectives from Different Disciplines*, ed. by M. G. Ash and T. Sturm (Mahwah, NJ: Lawrence Erlbaum, 2005), pp. 305–342.

chapter 1 by Lefkaditiou examines the local standardization attempts, and chapter 9 by Paulsen, the test's 'shelf life'. Thereby, the contributors highlight the elusive and chameleonic nature of the test coming in many forms and versions,<sup>8</sup> exploring its materiality and commercialization.

3) The historical training of previous researchers led them to develop a contextual history of what they called 'mental measurement', notably developed by Sokal, Zenderland, Carson and others, usually located in French, and/or North-American society.<sup>9</sup> Such narratives were mainly focused on the work of physicians and psychologists. It led to questions about what happened in all the other contexts in which IQ testing was introduced.<sup>10</sup> Thus, a third shift moved historians' interests towards the tests and the testing conducted by diverse professionals in different parts of the world. This opening-up has certainly enriched our historical knowledge.

4) A fourth shift came, as a side effect of the previous turns. It refers to the professionals conducting the historical research. Traditional research on historical figures and topics related to psychology had mainly been conducted by psychologists themselves. Historians of science would not care about it until the last decades of the twentieth century (see shift 1). At that time, some historians of science would start to discover their interest in this area, with the history of IQ testing quickly becoming a central point of interest.

In the last two decades, an array of professionals working on "Psy-history" has further expanded, with historians of education, cultural historians, Science and Technology Studies scholars, anthropologists, general historians, media studies experts, legal experts, archivists, amateur historians, and many more getting on board. Independently,

---

8 See also the personification of the 'big family' metaphor used by the editors in the introduction.

9 Mikel Sokal (ed.), *Psychological Testing and American Society (1890–1930)* (New Brunswick, NJ: Rutgers University Press, 1987); Leila Zenderland, *Measuring Minds: Henry Herbert Goddard and the Origins of American Intelligence Testing* (Cambridge, UK: Cambridge University Press, 1998) and John Carson, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics* (Princeton, NJ: Princeton University Press, 2007).

10 Annette Mülberger, 'The need for contextual approaches to the history of mental testing', *History of Psychology*, 17 (2014), 177–186.

whether we like this development or not, it is important to be aware that it introduces a greater variety, with each history carrying the imprint of the author's professional training and specialization.

Thus, we can conclude that the book is timely in several ways. Its content echoes the cacophony of approaches and interpretations we have today. It is a book without the collaboration of a psychologist, offering something like a 'strangeness-perspective' in the anthropological sense<sup>11</sup>, dominated by a perceivable 'puzzled amazingness' about the intelligence test's display of power in the 'cold' Norwegian social environment.

## Bibliography

- Boring, Edwin (Garrigues), *A History of Experimental Psychology* (New York: Appleton Century-Crofts, 1950 [1929]).
- Carson, John, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics* (Princeton, NJ: Princeton University Press, 2007). <https://doi.org/10.1515/9780691187679>
- Gigerenzer, Gerd and Sturm, Thomas, 'Tools = Theories = Data? On some circular dynamics in cognitive science', in *Psychology's Territories: Historical and Contemporary Perspectives from Different Disciplines*, ed. by M. G. Ash and T. Sturm (Mahwah, NJ: Lawrence Erlbaum, 2005), pp. 305–342.
- Gundlach, Horst, 'The Hipp chronoscope as totem pole and the formation of a new tribe', *Teorie & Modelli*, 1.1 (1996), 65–85.
- Mülberger, Annette, 'The need for contextual approaches to the history of mental testing', *History of Psychology*, 17 (2014), 177–186. <https://doi.org/10.1037/a0037487>
- Mülberger, Annette, 'Appropriation of psychological testing in the Spanish pedagogical context', in *The Circulation of Science and Technology: Proceedings of the 4th International Conference of the ESHS*, by A. Roca-Rosell (Barcelona: SCHCT-IEC, 2010), p. 626.
- Mulder, Ernst and Heyting, Frieda, 'The Dutch Curve: The introduction and reception of intelligence testing in the Netherlands, 1908–1940', *Journal of the History of the Behavioral Sciences*, 24.4 (1998), 349–366.
- Rheinberger, Hans Jörg, *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube* (Stanford, CA: Stanford University Press, 1997).

---

11 Steve Woolgar and Bruno Latour, *Laboratory Life: The Construction of Scientific Facts* (Princeton, New Jersey: Princeton University Press, 1986).

Sokal, Mikel (ed.), *Psychological Testing and American Society (1890–1930)* (New Brunswick, NJ: Rutgers University Press, 1987).

Woolgar, Steve and Latour, Bruno, *Laboratory Life: The Construction of Scientific Facts* (Princeton, New Jersey: Princeton University Press, 1986).

Teigen, Karl Harvor, 'A note on a Norwegian precursor of Charles Spearman', *Journal of the History of the Behavioral Sciences*, 22.1 (1986), 61–63.

Teigen, Karl Harvor, *En Psykologihistorie* (Oslo: Fagbokforlaget, 2015 [2004]).

Zenderland, Leila, *Measuring Minds: Henry Herbert Goddard and the Origins of American Intelligence Testing* (Cambridge, UK: Cambridge University Press, 1998).





# Index

---

- Aagaard, Knut Erik 227, 331–333, 450  
Aall, Anathon 42–44, 166–168, 182  
Aas, Ingeborg 116, 342–343, 356  
adaptation of tests xi, 3, 5–6, 9–10,  
13, 19–21, 23, 30–32, 41, 50, 55–56,  
60, 65–66, 73, 87, 136, 155, 176,  
200, 223, 226, 230, 233, 237, 248,  
269–274, 279–282, 284–287, 289,  
291, 295, 297, 303–308, 310, 312,  
316, 318, 321–322, 325, 327, 329,  
332, 334, 340, 342–343, 357, 364,  
369, 376, 380, 384, 388, 404, 435,  
445–446, 450, 456, 460  
administration of tests xiv, 56, 84,  
200, 221, 224, 236, 245, 249, 251,  
273, 275, 293, 316, 340, 348, 408,  
415–417, 420, 453, 455  
African Americans 306–307, 408, 416,  
426, 429  
Aker 62–64, 191, 225  
Arnesen, Harald 114–115, 183, 279  
Ås, Arvid 148–149, 349  
Bergen 13, 34, 37–39, 125, 167, 189,  
191–193, 196–197, 199–200, 202,  
204–209, 211, 243, 248, 263, 280,  
289–291, 295, 309, 339, 402, 448  
Binet, Alfred 4, 5, 6, 15, 19, 30, 60,  
101, 113, 119, 269, 271, 274, 406,  
411, 424, 425. *See also* Binet-Simon  
test; Stanford-Binet test  
Binet-Simon test 4–7, 13, 30–31, 36,  
38–42, 46, 49, 56–57, 73, 85, 87,  
110, 112, 114, 120, 176, 183, 215,  
228, 270, 275, 292, 406–411, 413,  
420, 425  
Blair, Tony 384–386  
Brigham, Carl 416–417, 426  
*Brown v. Board Supreme Court* 10,  
428–429, 431, 433  
Buros, Oscar Krisen 276  
Burt, Cyril 5, 60, 435  
Cappelen, J. W. 25, 269, 271, 281,  
283–284, 298, 383, 385, 387, 455  
Carson, John 5, 7, 9–10, 31, 38,  
102, 105, 120, 160–161, 244–245,  
272, 274, 338, 401–403, 406–407,  
410–413, 415–417, 424, 426, 463  
Cattell, James McKeen 320, 418  
Chapman, Paul Davis 10, 402–403,  
408–409, 412–413, 416–419  
class 11–12, 20, 23, 32, 36, 47, 63,  
113–114, 121, 140, 306–307, 325,  
328, 334, 340–341, 355–357, 405,  
409–410, 412, 418–419, 423–424,  
426, 428–437, 447–448, 451  
classism 23, 140, 402–403, 405,  
434–435, 446, 451  
Clinton, Bill 384  
Cold War 428–429  
conservative politics 36, 226, 339–340,  
347, 349–352, 357–358, 372–373,  
376–377, 382, 384, 409, 446  
culture 10, 20, 43, 110, 189, 271,  
304, 306–308, 323, 325, 327–329,  
331, 334, 382, 390, 409, 411, 430,  
448–450, 460  
'culture fair' test 323  
Dahlstrøm, Sigurd 40–41, 44, 49, 56,  
60, 62, 102, 109, 112, 460  
Dale, Erling Lars 363, 366, 369, 380  
Daston, Lorraine 3–4, 102, 258, 271

- Dedichen, Hans Henrik 146–147
- Denmark 9–10, 12, 14, 35, 49, 56, 65, 73–74, 76–79, 81–84, 87–89, 92–95, 106, 132–133, 137, 139, 175, 190, 193, 221, 244, 318, 356, 391, 394, 402, 404–405, 417, 422, 444, 459
- Dewey, John 366–367
- educational psychology 5–6, 17–18, 20–21, 43, 45, 87, 189–193, 195–196, 199–202, 204–212, 215–217, 219–221, 223–239, 281, 292, 303, 309–313, 316, 325, 328–331, 420, 431, 444, 450
- Eide, Kjell 378–380, 383
- Elgenström, Signe 77, 83, 87, 91–92
- Eng, Helga 16, 62, 85–86, 89, 92, 160, 172, 174–185, 276–282, 285–286, 288–289, 298, 364–367, 375, 390, 394, 432
- Enlightenment, the 32, 107
- epistemic infrastructures 261–262, 265
- Erlander, Tage 372–373
- ethnicity 20, 328, 405, 424, 429, 447–449
- eugenicist 30, 43, 45, 49, 410, 448
- eugenics 4, 8, 11–12, 15–16, 31, 37, 42, 45, 54, 65, 115, 131–134, 136–137, 139–143, 150, 154–157, 190, 305, 330, 405, 410–411, 416, 448–449
- Europe 11–12, 32, 41, 43, 51, 73, 104, 110, 132, 137, 160, 162, 175, 184–185, 223, 274, 385, 408, 449, 459
- Evang, Karl 138–141, 143, 150–152
- Faye-Hansen, Elisabeth 75, 84–85
- 'feeble-mindedness' 5, 30, 33–35, 38–39, 42, 53, 59, 63–65, 82, 87, 104, 113–114, 121, 135–136, 140, 147, 154, 156, 294, 341, 403, 406–412, 416, 419, 424, 427
- Fish Intervention Studies (FINS) 247–248, 250–251, 258
- FINS-KIDS 248, 250–251, 258
- folkeskole* (the people's school) 74, 80, 341–342, 345, 355–356
- forensic psychiatry xi, 15, 37, 45–46, 103, 110–112, 114, 118–119, 121, 123, 127, 452
- Foucault, Michel
- Foucauldian 8
- France 4–7, 9, 30–31, 43, 73, 102, 105, 160, 215, 244, 269, 272, 274–275, 284, 297, 388, 402, 404, 406, 461, 463
- Froestad, Jan 12, 32, 35, 190–191, 193, 195, 210, 291, 294, 342
- Galison, Peter 102, 258, 271, 290
- Galton, Francis 4–5, 47, 413
- Gerhardsen, Einar 165–166, 185
- Germany 5, 41, 79, 85, 108, 132–133, 136, 144, 151–152, 160–161, 169–171, 174–176, 277–279, 343, 352, 355, 358–359, 364, 366, 385, 390, 412, 417
- Gjessing, Hans Jørgen 191, 200, 208–211, 217, 309
- Goddard, Henry Herbert 5, 10, 30, 43, 274–275, 402, 406–412, 416, 419, 463
- Grennes, Otto 36, 51, 83, 90–91
- Greve, Ellen 208
- Hagtvet, Knut A. 321–324
- Hall, Granville Stanley 43, 45, 47
- Hansen, Klaus 169–170, 181
- Haraway, Donna 258–259
- Haug, Peder 88, 191, 212
- Havin, Henry 182–184
- historicization 1, 217, 259, 405, 459, 461–462
- Houghton Mifflin 46, 59, 269, 271, 273, 275, 282–283, 287, 292, 296, 298, 413
- intellectual disability 2, 15, 30, 33, 35, 42, 49, 53, 101, 103–104, 112, 114, 122, 124, 126, 131, 136, 150, 154–155, 244, 294, 303, 310, 313, 447
- mental disability 2, 16, 35, 125, 156, 459

- International Classification of Diseases (ICD) 15, 122, 124, 127, 296
- Jaederholm, Gustav 83, 86, 88–90, 92
- Jensen, Arthur 305, 328
- Jung, Carl Gustav 365
- Kaufman, Alan S. 305–306, 308, 311, 328–329, 332, 335
- Kayser, Hans Georg 328–329, 332
- Keilhau, Wilhelm 169, 181
- Keller, Christian 49, 88, 425
- Kløve, H. 208, 295
- Knudsen, Gunnar 169–170
- Kven 318, 327, 405, 449
- labor market 161, 165, 169–173, 185–186, 434
- Labor Party 11, 21, 136, 161, 164–165, 169–171, 175, 180, 345–346, 352, 354–358, 363, 371–373, 376, 378, 380, 385–386, 393, 405, 430, 432, 446
- late maturation 141, 143, 149
- Latour, Bruno 78–79, 93, 464
- legal reforms 15, 34, 45, 102–103, 115–116, 119–120, 123, 125, 294
- disability rights 156
- Lipmann, Otto 364–366
- Lippmann, Walter 406, 425
- Lofthus, Johan 13, 31, 48–64, 77, 87–88, 92, 113, 132, 135, 144, 146, 279–280, 435
- Osloprøven* 13, 57, 197–198, 200
- Lønning, Per 346–347, 351–352, 359
- Looft, Carl 38–39, 102
- mangelfullt utviklede sjelsevner*  
[deficient mental capabilities]  
114, 133–135, 156
- McGraw-Hill 282, 428
- measuring instrument 1, 3, 5, 12, 17, 20–21, 23, 42, 83, 108, 128, 132, 162, 166, 217, 222, 224, 264, 309–310, 326, 343, 355, 369, 391–394, 404, 406, 413, 418, 431, 449, 451, 462
- Mercer, Jane 307–308, 325–328, 332
- Merrill, Maud 269, 271, 276–277, 284–288, 296–297
- Meyer, Henning 88
- Monrad Jacobsen, Julie 29–30, 33, 35–36, 41–42, 83–84, 89–90, 92
- Mülberger, Annette 7–10, 23, 30–31, 66, 73, 102, 215, 270, 338, 459, 461, 463
- Munch, Ole B. 146–148
- Münsterberg, Hugo 166, 169, 180, 187
- National Defense Education Act (NDEA, 1958) 428–431, 433–434
- National Quality Assessment System (NQAS) 387–388, 391–394
- Nazification 355, 359
- Nazism 133, 139, 150, 152, 160, 170
- Nordic 9, 14, 29, 33, 35, 74–80, 83–85, 87, 92–95, 132–133, 137, 165, 169, 190, 209, 216, 239, 244, 264, 337–339, 348, 351–352, 354, 357, 368, 373, 384, 404, 417, 426, 444–445
- normed/norming for tests 6, 13, 20, 223, 229–231, 264, 304, 306, 310–311, 316–319, 323, 325, 329, 332, 334, 406, 412, 420, 436, 445, 448, 450
- Norwegianization 318, 327, 405, 448–450
- objectivity 1–2, 12, 14–15, 36, 76, 83, 90, 102, 114, 120–122, 126–128, 139, 192, 200, 218, 232, 235, 237–238, 342–343, 354, 381, 405, 407, 410, 419, 423, 430, 432, 436–437, 447, 451–453, 459
- OECD (Organisation for Economic Co-operation and Development) 21, 363, 370, 378–379, 383–386, 388–389, 392–393
- Oftedal, Dagny 148, 339–340, 345, 355, 379, 385, 404
- Ording, Hans 186
- Ørum, Hans Peter 49, 83–84, 88–90
- Østby, Håkon 196–198, 296

- Pedagogisk Forskningsinstitut*  
[Pedagogical Research  
Department] 277, 298–299, 339,  
360, 363–364, 367, 369, 375–382,  
390, 392, 395, 455
- pedagogy 1, 9, 12, 14, 18–20, 35,  
41, 52, 54, 76, 78, 85–88, 90–91,  
106, 145, 167, 174–175, 183, 190,  
195–198, 201, 204–205, 208–211,  
218–221, 223–224, 226, 228,  
234–235, 244, 250, 261–262, 265,  
309–311, 327, 331–333, 338, 340,  
342–343, 345–346, 349–350,  
352–354, 356, 364–367, 369, 372,  
374–382, 389–394, 402, 406,  
421–422, 432, 446, 454–456, 461
- Pedersen, Marie 77, 81, 83, 87, 91–92,  
280
- Pihl, Joron 217, 329–333, 335, 450
- professional judgement 453
- Programme for International Student  
Assessment (PISA) 22, 363–364,  
370, 383–394
- Project Metropolitan 337, 339–340,  
352–353
- psychiatrists 12–14, 16, 18, 31, 40,  
48–50, 52, 54, 58, 61–62, 65, 76,  
79, 82, 84, 87–89, 91, 94, 101, 107,  
110, 112, 114, 118–120, 123, 127,  
131–132, 135–136, 139, 142–146,  
148–149, 151–152, 156–157, 184,  
193, 199, 215, 281, 406, 421, 423,  
435, 452–455, 460
- psychiatry 6, 10, 12–16, 18, 36–37,  
39–40, 45–46, 48–49, 51–52, 54–56,  
58–65, 76, 103, 107–122, 126–128,  
136, 144–149, 151, 156, 167, 193,  
195, 202, 225, 231–234, 238–239,  
244, 292, 294, 311, 402, 422, 446,  
452–453, 456. *See also* forensic  
psychiatry
- psychologists 3–6, 12–14, 18, 20, 30,  
35–37, 42–43, 50, 52, 58, 60, 62,  
65, 76, 79, 83–84, 86, 88, 91, 93–94,  
101, 148–149, 157, 160–161, 163,  
165–171, 174, 176, 178–182, 184–  
186, 192, 194–197, 200–207, 209,  
216, 219–221, 224, 230, 233–235,  
239, 249, 274, 276, 278–279, 288,  
290–294, 297, 305–306, 309, 311,  
316, 320, 328, 331, 340, 347–348,  
350, 352, 365, 369, 390, 402, 406,  
408, 412, 416, 418, 424–426, 431,  
445, 452–456, 459–460, 462–464
- psychology 3–10, 12, 16–17, 23–24,  
30–32, 36–38, 42–46, 51, 61, 65–66,  
73, 76, 78, 82, 85–87, 90–92, 102,  
104, 108–111, 123, 145, 147–149,  
155, 159–171, 174–190, 193, 194,  
196–197, 201–211, 216, 218–221,  
224–225, 229, 233, 235–237, 239,  
244, 255, 259, 261, 269–278, 281,  
290–296, 303, 305–309, 312–313,  
317–320, 332, 337–338, 341,  
344, 349–350, 352–354, 364–366,  
372–373, 377, 379, 389, 401–402,  
411, 415, 417–420, 426, 428, 431,  
444–445, 454–456, 459, 461–463.  
*See also* educational psychology
- Educational Psychological Service  
(EPS) 6, 18, 20–21, 189, 191–193,  
195–196, 199–202, 204–212,  
215–217, 219–221, 223–239, 303,  
309–313, 316, 325, 328–331, 444,  
450
- psychometrics 8, 12, 19, 22, 148, 175,  
237, 259, 312, 364–365, 367, 390,  
427
- psychotechnics 16–17, 85–87, 89, 159–  
163, 165–187, 277–279, 364–366
- Psychotechnical Institute 16, 87,  
160, 171–172, 176, 180, 182–184,  
277–279, 364–366
- race 11–12, 20–21, 23–25, 37–38, 45,  
51, 59, 133, 140, 173, 305–308, 317,  
326, 328, 405, 410, 413, 416–418,  
426–431, 433–437, 443, 447–450
- racism 8, 23, 45, 59, 140, 306, 330, 335,  
338, 357, 402–403, 405, 413, 426,  
434–435, 446, 449, 451
- anti-racist 306, 426
- Rakvåg, Gunnar 202–205
- Rasmussen, Augusta 62, 131, 135,  
143–148, 152, 280

- Ravneberg, Bodil 12–13, 32, 35,  
190–191, 193, 195, 210, 291, 294
- Reagan, Ronald 383–384
- Reymert, Martin L. 36, 42–49, 56–57,  
59–60, 167, 279, 422, 460
- Ribsskog, Bernhof 62, 84–85, 88, 92,  
280, 341–343, 356
- Rifbjerg, Sofie 77, 83, 87, 89
- Romani 448–449
- Sámi 318, 327, 405, 449–450
- Sandven, Johs 199–201, 206, 269,  
271, 281–283, 285–290, 292–293,  
295, 348–349, 367–370, 375–377,  
379–382, 388–390, 393–394, 432,  
435
- Sarsgaten school 36, 54, 63–64
- Schiøtz, Carl 48, 51–52, 54, 57
- Schjelderup, Harald 44, 46, 163, 165,  
169–170, 181, 186, 348–349
- school and educational reforms 17,  
21–22, 32, 38, 45, 54, 73, 78, 85, 91,  
95, 173, 189, 191, 193–194, 196,  
210–212, 346, 356–357, 373–376,  
378, 380, 386, 393, 414, 420, 422,  
429, 431–432, 435–436, 444, 446,  
453, 455, 460
- pedagogical reform 35, 175
- reform pedagogy 85, 446, 453, 455,  
460
- Schröder, Gerhard 385–386
- Seashore, Carl Emil 43, 45
- segregation 11–13, 17, 22, 35, 37, 44,  
49, 51, 55, 81, 154, 156, 173, 191,  
195, 294, 341–342, 409, 416, 429
- Jim Crow 173, 408, 416, 429
- Sethne, Anna 51, 84–85, 88
- sexual offences 115
- Simon, Théodore 4, 30, 41, 61, 113,  
119, 274, 275, 406. *See also* Binet-  
Simon test
- Sivertsen, Helge 352, 371, 373–375
- Skard, Åse Gruda 37, 47–48, 76,  
148–149, 165, 194
- Skjervheim, Hans 347, 350, 380–382
- slavery 11, 405, 434–435
- Smith-Hughes Act 414–415, 418–419,  
422, 433–434
- Sørensen, Marta 41, 83–84, 383, 432
- Spearmen, Charles 5, 320, 459
- special education 4, 14, 17, 20–23,  
29, 32–35, 37, 41, 55, 74–77, 79–82,  
91–94, 193–195, 201–204, 207, 217,  
226, 303, 307, 309, 313, 329–330,  
332, 341, 403–404, 419–421, 423,  
431–432, 434, 437, 446, 450, 459
- special schools 13–14, 17, 29–30,  
33–39, 41–43, 48–49, 52–55, 57–58,  
62–64, 74–85, 87, 91–94, 135, 148,  
193–195, 199, 203, 294, 310, 431,  
444
- Nordic Special School Association  
77–78, 83–84, 87, 93–94
- Oslo Special Schools Association  
36, 38, 42
- standardization of tests 3–5, 9,  
13–14, 19–20, 30–32, 36–37, 41,  
44–48, 56–57, 61–62, 64–66, 73, 85,  
87–88, 196, 200, 208–209, 211, 218,  
228–229, 232, 234–235, 237, 239,  
243–244, 248–250, 255, 264, 271,  
273, 279–280, 282, 284, 286, 290,  
297, 303–304, 310, 312, 317–318,  
320–321, 324, 326, 332, 340, 383,  
417–418, 420–422, 431, 436–437,  
445, 455, 460, 463
- Stanford–Binet test xi, 5–6, 13, 29–31,  
40, 46–48, 50, 56–57, 59, 61–62,  
87–88, 132, 215, 269–276, 278–280,  
283–284, 289, 296–297, 310–312,  
369, 375, 412–413, 416–418,  
420–421, 431, 448, 455, 460
- sterilization 2, 15–16, 45, 62, 115, 118,  
131–145, 147–157, 330, 410, 446,  
448, 453, 460
- Norwegian Sterilization Act of 1934  
15, 131–133, 135–137, 143–144, 147,  
149–150, 152–153, 156
- Penal Code Committee 132–133,  
135–136
- Sterilization Board 16, 62, 131–132,  
134–135, 138–139, 141–143, 145,  
152, 453

- Stern, William 5, 85, 117, 161, 364–366, 412
- Sverdrup Lunden, Turi 281, 288–291, 293–295
- Sweden 9–10, 12, 14, 33, 35, 49, 65, 73–79, 82–84, 87–89, 92–95, 132–133, 137, 139, 193, 244, 287, 354–355, 371–376, 379, 394, 402, 404, 422, 459
- Tamm, Alfchild 49, 88
- Taylorism 161, 171, 414
- Terman, Lewis 5, 10, 30–31, 43, 46, 52, 57–61, 117, 119, 269, 271, 274–276, 279, 282, 284–287, 292, 402, 412–420, 425–427, 448
- Terman–Merrill test 19, 197–198, 269–272, 274, 278, 280, 282, 284, 288–289, 291–293, 295–297, 310, 375, 450
- Thatcher, Margaret 383–384
- Torshov school 63
- translation xi, 3, 5–6, 10, 13, 19, 24–25, 29, 31, 38, 40, 46–47, 49, 61, 73, 80–81, 85–86, 104, 106–107, 125, 134, 150, 165, 179, 215, 218–223, 228–231, 251, 254, 261, 264, 269, 272, 274, 278–286, 289, 297, 303–304, 316, 318, 321, 366, 389, 406, 408, 420, 436, 448
- Trondheim 34, 77, 85, 280, 291, 293–294, 312–313
- Ullevålveien school 29, 36, 54, 62–64
- unaccountability 15, 101–102, 104–105, 108, 115–117, 119–127, 134, 447, 452
- Undheim, Johan Olav 312–313, 315–316, 319–326
- USA 5–12, 20–23, 25, 30–31, 33, 41, 43, 45–47, 60, 73, 77, 86, 102, 132, 146, 150, 160–162, 164, 167, 169, 173, 176, 180, 215, 229, 244, 246, 248, 254, 259, 261, 264, 269, 271–272, 274–276, 279–280, 282–284, 286, 292, 294–295, 297, 303–307, 312, 316–318, 320–321, 323–324, 326, 328, 333–334, 338–339, 357, 366–367, 369, 379, 383–384, 390, 401–406, 408–412, 414–419, 421–430, 432–437, 444–446, 448–450, 461, 463
- Vismann, Cornelia 272
- Vogt, Ragnar xi, 36–37, 42, 44–46, 48–49, 51, 61, 110–112, 116, 118, 133, 136, 295
- Vormeland, Oddvar 191, 349–350
- Waal, Nic 142–143, 151–152, 281
- WAIS (Wechsler Adult Intelligence Scale) 3, 124, 218, 455
- Wallin, J. E. Wallace 424–425
- Wechsler, David 6, 218, 236, 305, 311, 313, 315–317, 323, 416, 445
- Wesman, Alexander G. 305–308, 311
- Whipple, Guy 416–417, 425–426
- Williams, Robert L. 6, 218, 274, 306–308, 317, 325
- Winge, Paul 110–112, 114
- WISC (Wechsler Intelligence Scale for Children) tests 2–4, 6–7, 18–19, 216–225, 227, 229–235, 237–239, 243–246, 248, 251, 254, 257–261, 263–266, 303, 305–306, 308, 310–313, 315–317, 319, 332–333, 437, 445, 454–456
- WISC-III 2–4, 305–306, 331, 445
- WISC-IV 219, 445
- WISC-R 20–21, 303–325, 327–334, 445, 450
- WISC-V 216–217, 219, 221–222, 239, 445, 454
- WPPSI-III (Wechsler Preschool and Primary Scale of Intelligence – Third edition) 243–248
- WPPSI (Wechsler Preschool and Primary Scale of Intelligence) 3, 221, 243–249, 252–253, 256–257, 259, 263–264, 266, 456
- World Book Company 275, 279, 282
- World War One 161–162, 304, 402,

- 415–417, 422, 426  
World War Two 6, 12, 21, 23, 132, 165,  
186, 191, 193, 281, 355, 367, 371,  
390, 393, 402, 422–423, 427–428,  
430–433, 445, 449  
Yerkes, Robert 261, 415, 417, 419, 426  
Zenderland, Leila 10, 30, 274,  
402–403, 406–411, 415–416, 424,  
426, 463





# About the Team

Alessandra Tosi was the managing editor for this book.

Lucy Barnes and Hannah Shakespeare proof-read this manuscript. Lucy produced the index.

Jeevanjot Kaur Nagpal designed the cover. The cover was produced in InDesign using the Fontin font.

Jeremy Bowman typeset the book in InDesign. The main text font is Tex Gyre Pagella and the heading font is Noto Serif.

Jeremy also produced the PDF, paperback, and hardback editions and created the EPUB. The conversion to the HTML edition was performed with epublius, an opensource software which is freely available on our GitHub page at [https:// github.com/OpenBookPublishers](https://github.com/OpenBookPublishers)

Lucy Barnes was in charge of marketing.

This book was peer-reviewed by two anonymous referees. Experts in their field, these readers give their time freely to help ensure the academic rigour of our books. We are grateful for their generous and invaluable contributions.



# This book need not end here...

## Share

All our books — including the one you have just read — are free to access online so that students, researchers and members of the public who can't afford a printed edition will have access to the same ideas. This title will be accessed online by hundreds of readers each month across the globe: why not share the link so that someone you know is one of them?

This book and additional content is available at  
<https://doi.org/10.11647/OBP.0471>

## Donate

Open Book Publishers is an award-winning, scholar-led, not-for-profit press making knowledge freely available one book at a time. We don't charge authors to publish with us: instead, our work is supported by our library members and by donations from people who believe that research shouldn't be locked behind paywalls.

Join the effort to free knowledge by supporting us at  
<https://www.openbookpublishers.com/support-us>

We invite you to connect with us on our socials!



BLUESKY  
[@openbookpublish](https://bsky.social)  
[.bsky.social](https://bsky.social)



MASTODON  
[@OpenBookPublish](https://hcommons.social)  
[@hcommons.social](https://hcommons.social)



LINKEDIN  
[open-book-](https://open-book-publishers)  
[publishers](https://open-book-publishers)



INSTAGRAM  
[openbookpublishers\\_](https://openbookpublishers_)

Read more at the Open Book Publishers Blog

<https://blogs.openbookpublishers.com>

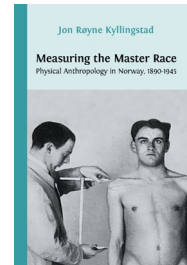
You may also be interested in...

## Measuring the Master Race

Physical Anthropology in Norway 1890-1945

*Jon Røyne Killingstad*

<https://doi.org/10.11647/OBP.0051>

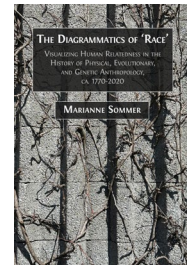


## The Diagrammatics of 'Race'

Visualizing Human Relatedness in the History of  
Physical, Evolutionary, and Genetic Anthropology  
ca. 1770-2020

*Marianne Sommer*

<https://doi.org/10.11647/OBP.0396>

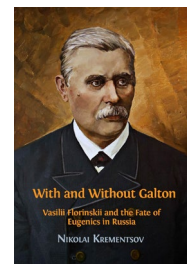


## With and Without Galton

Vasilii Florinskii and the Fate of Eugenics in Russia

*Nikolai Krementsov*

<https://doi.org/10.11647/OBP.0144>

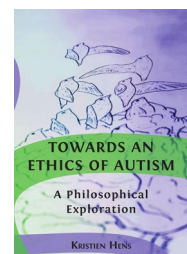


## Towards an Ethics of Autism

A Philosophical Exploration

*Kristien Hens*

<https://doi.org/10.11647/OBP.0261>



# HISTORICIZING IQ TESTING

## Intelligence Assessments and their Role in Norwegian Society from the 1900s to the Present

EDITED BY HÅKON AAMOT CASPERSEN AND JON RØYNE KYLLINGSTAD

*Between the turn of the previous century and today, questions about human intelligence in Norway have played out in many of the same ways they have in the rest of the world. IQ has been put to use for good and ill, helping children in school and supporting horrific eugenic programs. This book follows the scientific history of Norwegian intelligence testing in fascinating detail. I learned a great deal.*

Prof Eric Turkheimer, University of Virginia

Intelligence testing has shaped modern society in profound ways, influencing education, psychology, law, and governance. This volume offers the first comprehensive study of the history of IQ testing in a Nordic country, shedding new light on its development, adaptation, and societal impact in Norway.

By tracing the evolution of intelligence tests—from their role in schools and special education to forensic psychiatry and criminal law—the book uncovers the tensions surrounding their use. Are these tests instruments of empowerment or tools of control? How have they shaped access to education, healthcare, and legal rights?

A key focus of this study is the transnational movement of intelligence tests, particularly between Norway, the USA, and other Nordic nations. It explores how tests have been translated, adapted, standardized, and used, raising questions about their claims to measure universal intelligence.

This volume challenges assumptions about IQ testing, placing practices of testing and the tests themselves at the center of historical analysis. By examining the Norwegian case, it contributes fresh insights to international scholarship, offering a vital perspective on the global history of intelligence measurement. Essential reading for historians, psychologists, and educators, this book redefines our understanding of intelligence testing in a changing world.

This is the author-approved edition of this Open Access title. As with all Open Book publications, this entire book is available to download for free on the publisher's website. Printed and digital editions, together with supplementary digital material, can also be found at <http://www.openbookpublishers.com>.

Cover photo © Helge Brekke / MUV, UiO, CC BY-SA 4.0

Cover design: Jeevanjot Kaur Nagpal

